

UNIVERSIDADE ESTADUAL DE CAMPINAS

Flávia Raquel Gonçalves Carneiro

**ANÁLISES ESTRUTURAIS E ESTUDOS DAS
INTERAÇÕES DAS PROTEÍNAS INT6 E NY-REN-21**

Tese apresentada ao Instituto de
Biologia para obtenção do Título de
Doutor em Genética e Biologia
Molecular na área de Genética
Animal e Evolução.

Orientador: Prof. Dr. Nilson Ivo Tonin Zanchin

Campinas
2006

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP**

C215a	Carneiro, Flávia Raquel Gonçalves Análises estruturais e estudos das interações das proteínas INT6 e NY-REN-21 / Flávia Raquel Gonçalves Carneiro. -- Campinas, SP: [s.n.], 2006. Orientador: Nilson Ivo Tonin Zanchin. Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia. 1. INT6. 2. NY-REN-21. 3. Duplo-híbrido de levedura. 4. Dicroísmo circular. 5. Transcrição genética. I. Zanchin, Nilson Ivo Tonin. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título. (rcdt/ib)
--------------	---

Título em inglês: Structural analyses and interaction studies of the INT6 and NY-REN-21 proteins.

Palavras-chave em inglês: INT6; NY-REN-21; Yeast two-hybrid system; Circular dichroism; Genetic transcription.

Área de concentração: Genética Animal e Evolução.

Titulação: Doutora em Genética e Biologia Molecular.

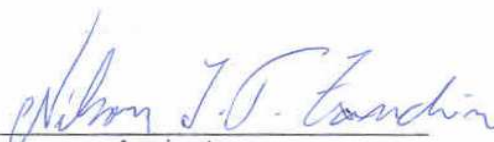
Banca examinadora: Nilson Ivo Tonin Zanchin, Anete Pereira de Souza, Celso Eduardo Benedeti, Beatriz Amaral de Castilho, José Camillo Novello.

Data da defesa: 30/05/2006.

Campinas, 30 de maio de 2006.

Banca Examinadora

Prof. Dr. Nilson Ivo Tonin Zanchin (Orientador)


Assinatura

Profa. Dra. Anete Pereira de Souza


Assinatura

Prof. Dr. Celso Eduardo Benedetti


Assinatura

Profa. Dra. Beatriz Amaral de Castilho


Assinatura

Prof. Dr. José Camillo Novello


Assinatura

Profa. Dra. Beatriz Gomes Guimarães

Assinatura

Prof. Dr. Roger Frigério Castilho

Assinatura

Prof. Dr. Marcelo Menossi Teixeira

Assinatura

AGRADECIMENTOS

Aos meus pais, Celia e Marcio, por todo carinho e dedicação, me ajudando sempre em todas as dificuldades surgidas e acreditando que eu poderia superá-las.

Ao Nilson pela orientação e conhecimentos adquiridos durante o desenvolvimento deste projeto.

Às minhas irmãs, Carina e Gabriella, e avós, Lourdes e Rosa, que sempre estiveram ao meu lado e torcendo por mim.

Às minhas grandes amigas Ana Cláudia, Daniela Ierardi, Daniela Soriano, Flávia Nery, Rayne e Lavínia por todos os momentos de diversão, apoio e companheirismo durante a realização deste trabalho.

Ao Paulo pelo carinho e paciência.

À Sandra e à Thais que foram minhas amigas dentro e fora do laboratório.

Às minhas colegas do laboratório Celisa, Patrícia, Juliana Smetana e Juliana Oliveira e ao Cédric pela amizade, ótimo convívio e sugestões para o desenvolvimento deste trabalho.

Ao Júlio pela amizade e ajuda durante os anos que convivemos no LNLS.

Às minhas amigas do Rio de Janeiro, Sara, Gabi, Ceíça e Daniela Franklin, que apesar da distância, sempre me apoiaram.

Um agradecimento especial à Tereza e Adriana que contribuíram com muita competência na parte experimental deste projeto.

À Kelly e ao Alexandre pelos momentos de descontração no laboratório.

Às funcionárias Givanil, Zildene e Luciana que sempre foram muito gentis e prestativas.

Ao Dr. Fábio Cesar Gozzo e ao Amadeu pelas análises de espectrometria de massa.

Ao Dr. Carlos Henrique Ramos e Dr. João Alexandre Barbosa pela participação e sugestões na banca prévia.

A todos os pesquisadores, funcionários e alunos do LNLS pelo convívio.

Às Profas Dras. Beatriz Castilho, Anete Pereira de Souza e Beatriz Gomes Guimarães e aos Profs Drs. Celso Eduardo Benedetti, José Camillo Novello, Marcelo

Menossi Teixeira e Roger Frigério Castilho por aceitarem participar da avaliação deste trabalho.

À FAPESP pelo apoio financeiro

Ao LNLS pela infra-estrutura fornecida e à UNICAMP pela oportunidade de aprendizado em diversas áreas da biologia.

A todos que direta ou indiretamente contribuíram para a realização deste trabalho.

ÍNDICE

AGRADECIMENTOS	iv
LISTA DE FIGURAS	viii
LISTA DE ABREVIACÕES	ix
RESUMO	xi
ABSTRACT	xii
1-INTRODUÇÃO	1
1.1- A proteína INT6	1
1.1.1- Identificação da proteína	1
1.1.2- Localização celular	5
1.1.3- A conservação da proteína INT6 e estudos funcionais	5
1.2- Identificação de ligantes protéicos da proteína INT6	8
1.3- A proteína NY-REN-21	12
1.3.1- Promotores e maquinaria de transcrição de mRNAs de eucariotos	12
1.3.2- Proteínas dedos de zinco tipo C ₄ e C ₂ H ₂	15
1.3.3- Domínios associados às proteínas dedos de zinco	17
1.3.3.1- Domínio KRAB	17
1.3.3.2- Domínio BTB/POZ	18
1.3.3.3- Domínio ZAD	19
1.3.3.4- Domínio SCAN	20
1.3.4- Identificação e aspectos da proteína NY-REN-21	21
2- OBJETIVOS	23
3- RESULTADOS	24
3.1- Resultados do rastreamento dos ligantes protéicos da hINT6	25
3.2- Artigo I	43
3.3- Artigo II	52
4- DISCUSSÃO	62
4.1- Triagem para a identificação dos ligantes protéicos da proteína INT6 humana	62
4.2- Análise da proteína INT6 de <i>A. thaliana</i> por dicroísmo circular (CD)	67
4.3- Análises da proteína NY-REN-21	71
4.3.1- Proteína NY-REN-21: purificação e dicroísmo circular	72
4.3.2- Análise da região central desordenada da proteína NY-REN-21 (<i>DCR</i> -	

<i>Disordered Central Region)</i>	73
4.3.3- Análise da região de dedos de zinco (<i>ZFR- Zinc Finger Region</i>)	77
4.3.4- SCAN: Purificação e análise por dicroísmo circular	77
4.3.5- O domínio SCAN da proteína NY-REN-21 pode se associar em homodímeros e formar heterodímeros com a proteína SCAND1	79
5- CONCLUSÕES	83
6- PERSPECTIVAS	85
7-REFERÊNCIAS BIBLIOGRÁFICAS	86
8- ANEXOS	101
8.1- ANEXO I	101
8.2- ANEXO II	113

LISTA DE FIGURAS

Figura 1.1: Esquema representativo da proteína hINT6 e suas formas truncadas hINT6T22 e hINT6T1139.	2
Figura 1.2: Modelo da iniciação da tradução em eucariotos	4
Figura 1.3: Diagrama representando o método do duplo-híbrido utilizado para isolar proteínas que interagem com a hINT6, sua forma truncada hINT6T22 e domínio PCI.	11
Figura 1.4: Formação do complexo de iniciação da transcrição pela RNA polimerase II.	14
Figura 1.5: Estrutura de regiões de interação ao DNA tipo dedo de zinco.	17
Figura 1.6: Representação da proteína NY-REN-21, seus domínios e regiões clonados para expressão, purificação e caracterização.	22
Figura 4.1: Alinhamento da sequência de aminoácidos das proteínas hINT6 e AtINT6.	71
Figura 4.2: Predição de regiões desordenadas na proteína NY-REN-21 pelo programa PONDR.	74

LISTA DE ABREVIACES

3-AT: 3-aminotriazol

AtINT6: protena INT6 de *Arabidopsis thaliana*

BTB/POZ: Domnio conservado encontrado em protenas dedo de zinco (*Broad-Complex*, *Tramtrack*, *Bric-a-brac/Poxvirus*, *Zinc finger*)

C₂H₂: Dedo de zinco formado por um on de zinco ligado a duas cistenas e duas histidinas

C₄: Dedo de zinco formado por um on de zinco ligado a quatro cistenas

CD: Dicrosmo circular (*Circular Dichroism*)

CSN: complexo COP9/signalossomo (*COP9 signalosome complex*)

DCR- Regio central desordenada da protena NY-REN-21 (*Disordered Central Region*).

EDTA: cido etilenodiaminotetractico

eIF: fator eucaritico de iniciao da traduo (*Eukaryotic translation Initiation Factor*)

hINT6: protena INT6 humana

hINT6T22: protena truncada que pode resultar da insero do genoma do MMTV no intron 9 do gene da INT6

hINT6T1139: protena truncada que pode resultar da insero do genoma do MMTV no intron 5 do gene da INT6

HTLV-1: vrus tipo I da leucemia humana de clulas T (*Human T Cell Leukemia Virus-1*)

INT: stio de integrao do MMTV

KRAB: Domnio conservado encontrado em protenas dedo de zinco (*Krppel-associated Box*)

MMTV: Vrus de tumor de glndula mamria de camundongo (*Mouse Mammary Tumor Virus*)

Moe1: *Microtubule over extended*

NES: sinal de exportao nuclear (*Nuclear Export Signal*)

NLS: sinal de localizao nuclear (*Nuclear Localization Signal*)

NY-REN-21: Antgeno encontrado em pacientes com carcinoma de clulas renais

PCI: Domínio de interação entre proteínas (Proteassomo, complexo COP9/signalossomo, fator 3 de iniciação da tradução).

PLZF: proteína dedo de zinco de leucemia promielocítica (Promyelocytic Leukemia Zinc Finger)

PONDR: Servidor para predição de regiões e proteínas naturalmente desordenadas (Predictors of Natural Disordered Regions)

PPARs: receptores ativadores de proliferação de peroxissomos (Peroxisome Proliferator-Activated Receptors)

PSIPRED: Servidor de predição de estrutura secundária (Protein Structure Prediction Server)

RRM: Motivo de reconhecimento de RNA (Recognition RNA Motif)

SCAN: domínio de dimerização (SRE-ZBP, CTfin-51, AW-1, Number 18)

SCAND1: *SCAN Domain containig 1*

SDS-PAGE: eletroforese em gel de poliacrilamida em presença de duodecil-sulfato de sódio (*sodium duodecyl-sulphate poliacrilamide gel electrophoresis*)

SEREX: Análise sorológica de bibliotecas de cDNA expressas (*serological analysis of recombinat cDNA expression libraries*)

T22: Tumor 22 encontrado em camundongo resultado da inserção do genoma do MMTV no intron 9 do gene INT6

T1139: Tumor 1139 encontrado em camundongo resultado da inserção do genoma do MMTV no intron 5 do gene INT6

TBP: proteína ligadora da sequência TATA (TATA Binding Protein)

TEV: *tobacco etch virus*

TF: fator de transcrição (Transcriptional Factor)

Tm: Temperatura média de transição

X-gal : 5-bromo-4-cloro- β -D-galactopiranosídeo

YNB: meio de cultura mínimo para crescimento de levedura (Yeast Nitrogen Base)

ZAD: Domínio conservado encontrado em proteínas dedo de zinco, restrito aos insetos (Zinc Finger Associated C4DM)

ZFP38: proteína dedo de zinco 38 de camundongo

ZNF: proteína dedo de zinco

RESUMO

O gene que codifica a proteína INT6 corresponde a um dos sítios de inserção do vírus de tumor de glândula mamária em camundongo (MMTV). Esta inserção pode levar à formação de proteínas truncadas sem a porção C-terminal e sem o domínio PCI, descrito como domínio de interação entre proteínas. Neste trabalho, realizamos 3 triagens para a identificação dos ligantes protéicos da proteína humana INT6 (hINT6) pelo método do duplo-híbrido de levedura. Embora interações específicas tenham sido identificadas, não foi possível a confirmação *in vitro* das novas interações isoladas. Para análises estruturais, usamos a proteína INT6 de *Arabidopsis thaliana* (AtINT6), visto que a proteína humana expressou em níveis muito baixos na fração solúvel do extrato de *Escherichia coli*. Análises de dicroísmo circular revelaram que a proteína AtINT6 é rica em estrutura do tipo α -hélice. A região que compreende os aminoácidos 172 a 415, incluindo o domínio PCI, foi identificada por proteólise parcial e espectrometria de massas como um domínio estrutural que após sua clonagem apresentou alto nível de expressão, solubilidade e estabilidade. Este trabalho envolveu também a caracterização da NY-REN-21, que foi isolada pelo duplo-híbrido para a identificação dos ligantes protéicos da proteína hINT6 e representa uma possível ortóloga para o fator de transcrição ZFP38 de camundongo. Ambas as proteínas apresentam um domínio de dimerização (SCAN), 7 domínios dedos de zinco tipo C_2H_2 na porção C-terminal e uma região central predita como desestruturada. A proteína NY-REN-21 se mostrou parcialmente desenhovelada e sua estrutura secundária é afetada pela incubação com EDTA. Ensaio de proteólise limitada, dicroísmo circular e fluorescência confirmaram que sua região central é intrinsecamente desordenada, além de apresentar mobilidade anômala em gel de SDS-PAGE e representa uma fração flexível da proteína. Não foi possível a caracterização da região dos dedos de zinco, pois esta região se mostrou altamente instável. O domínio SCAN da NY-REN-21 é capaz de formar homodímeros e heterodímeros com a proteína SCAND1. Esta interação pode representar uma forma de regulação da atividade da proteína NY-REN-21.

ABSTRACT

The INT6 gene was reported as a frequent genome integration site of Mouse Mammary Tumor Virus (MMTV). This integration may result in truncated forms of INT6 protein lacking its C-terminal region and the PCI domain. It has been reported that this domain is involved in protein-protein interaction. We performed 3 yeast two-hybrid screens in order to identify the human INT6 (hINT6) interaction partners. Although two previously described specific interactions were identified in these screens, it was not possible to confirm the new interactions by *in vitro* binding assays. The *Arabidopsis thaliana* INT6 ortholog (AtINT6) was used for structural analyses, since the hINT6 was insoluble following expression in *Escherichia coli*. AtINT6 showed CD spectra with high helical content. The region comprising amino acids 172 to 415 forms a compact protease resistant domain as determined by limited proteolysis and mass spectrometry analyses. This domain, comprising the PCI domain sequence, was cloned and expressed in *E. coli* and showed high solubility and stability. This recombinant protein has the potential to serve as a model protein for three-dimensional structure determination of the PCI domain. We also studied the NY-REN-21 protein, which was isolated as a potential hINT6 interaction partner and represents a putative ortholog of the mouse ZFP38 transcriptional factor. Both proteins are C₂H₂ type multifinger proteins, containing a conserved oligomerization domain (SCAN) in the N-terminal region and a predicted disordered central region. Our analyses showed that full-length NY-REN-21 is partially unfolded and its secondary structure content is affected by incubation with EDTA. The central region of NY-REN-21 shows an aberrant mobility on SDS-PAGE and is intrinsically unstructured as revealed by circular dichroism, fluorescence and limited proteolysis. The zinc finger region was not characterized because of its unstable nature. The recombinant SCAN domain of NY-REN-21 can form homodimers and heterodimers with the SCAND1 protein. This interaction may represent a novel regulatory mechanism of NY-REN-21 activity.

1-INTRODUÇÃO

1.1- A proteína INT6

1.1.1- Identificação da proteína

A proteína INT6 foi identificada por 3 maneiras independentes. Inicialmente foi demonstrado que o gene INT6 é o sítio de inserção do vírus de tumor de glândula mamária em camundongo (MMTV, *Mouse Mammary Tumor Virus*, Marchetti *et al.*, 1995). Além disso, o cDNA que codifica a proteína INT6 humana (hINT6) foi isolado em uma triagem através do método do duplo-híbrido usando a oncoproteína Tax do vírus tipo I da leucemia humana de células T (HTLV-1) como isca (Desbois *et al.*, 1996). Finalmente, a proteína INT6 foi caracterizada como uma subunidade do fator 3 eucariótico de iniciação da tradução (eIF3, Asano *et al.*, 1997).

A inserção do MMTV interrompe o gene da INT6 no intron 5 ou 9, correspondendo ao tumor 1139 (T1139) e tumor 22 (T22), respectivamente, encontrados em camundongos. A direção da transcrição do genoma do vírus integrado está em orientação oposta a do gene da INT6, podendo resultar na produção de proteínas truncadas sem a região C-terminal. No T1139 a proteína truncada resultante (INT6T1139) está *in frame* com um códon de terminação na porção U3 da sequência LTR (*long terminal repeat*) do DNA viral. Nesta forma truncada apenas um aminoácido é adicionado à sequência normal da INT6, após o aminoácido da posição 317. Entretanto no T22, a proteína truncada resultante (INT6T22) é uma quimera composta da sequência normal de aminoácidos da INT6 ligada a uma nova sequência de aminoácidos codificada pelo intron 9 do gene *int6* e a sequência de nucleotídeos reversa da região LTR do MMTV (Marchetti *et al.*, 1995, Figura 1.1)

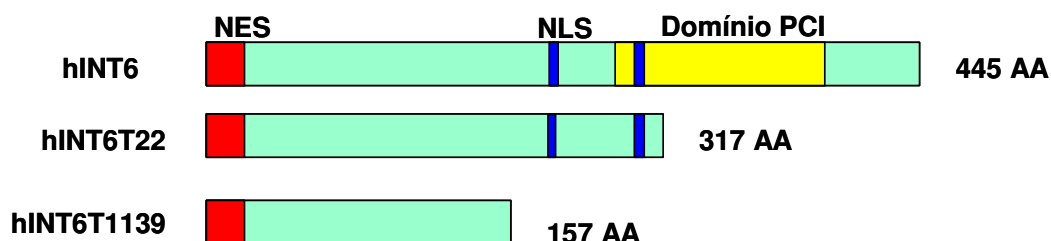


Figura 1.1: Esquema representativo da proteína hINT6 e suas formas truncadas hINT6T22 e hINT6T1139. NES: sinal de exportação nuclear (vermelho); NLS: sinal de localização nuclear (azul) e domínio PCI (Proteassomo, complexo COP9/sinalossomo, fator 3 de iniciação da tradução).

Alguns estudos foram realizados buscando entender se as formas truncadas da INT6 humana (hINT6) apresentavam ganho de função, se atuavam como mutantes dominantes negativos ou se acarretavam a perda de função da hINT6. Miyazaki *et al.* (1997) mostraram que a hINT6 é freqüentemente perdida em carcinomas primários de mama (30% de todos os pacientes examinados), sugerindo que a perda de função da hINT6 contribui para a transformação maligna. Buttitta *et al.* (2005) estudando tumores primários de pulmão verificaram que em 73% dos casos, os níveis de expressão do RNA mensageiro (mRNA) da hINT6 estavam aumentados e que em 27% havia baixa expressão dos transcritos de hINT6. No último caso os tumores se mostraram mais agressivos com redução não só do número de pacientes sobreviventes, assim como do tempo de sobrevida. Entretanto, dois outros estudos indicaram que a expressão da forma truncada hINT6T1139 pode levar à transformação maligna em células epiteliais de mamíferos (Rasmussen *et al.*, 2001) e células NIH3T3 (Mayeur & Hershey, 2002) mesmo na presença de 2 alelos funcionais do gene, sugerindo que esta proteína atue de maneira dominante.

A proteína Tax do HTLV-1 é um ativador da transcrição da expressão gênica celular e é requerida para a replicação viral *in vivo*. Tax regula positivamente a expressão de genes virais *in vitro* e também ativa a transcrição de muitos outros genes, incluindo interleucina 1 e seu receptor Bcl-xL, o antígeno nuclear de proliferação celular (PCNA), vimentina, fator de necrose tumoral- α (TNF- α), Krox 24,

interferon γ e o fator de crescimento transformante- β (TGF- β , Yoshida, 2001). Para identificar as proteínas capazes de interagir com a oncoproteína Tax do HTLV-1, foi realizada uma triagem pelo método do duplo-híbrido usando a proteína Tax como isca (Desbois *et al.*, 1996). Dois clones contendo a sequência da proteína INT6 foram isolados e a interação confirmada por duplo-híbrido realizado em células de mamíferos e por ensaios de co-imunoprecipitação.

A proteína INT6 também foi identificada como a subunidade de 48 kDa do eIF3. O eIF3 é o mais complexo fator de iniciação conhecido, tanto com respeito à função, quanto à sua composição. Foram identificados, até o momento, 12 subunidades para o fator de mamíferos sendo 6 conservadas em *Saccharomyces cerevisiae* (Hershey & Merrick, 2000; Browning *et al.*, 2001; Morris-Desbois *et al.*, 2001).

De acordo com o modelo do mecanismo de iniciação da tradução (Hershey & Merrick, 2000; Pestova *et al.*, 2001; Sonenberg & Dever, 2003), durante a etapa de iniciação, primeiramente as subunidades ribossomais são dissociadas em 40S e 60S. Os fatores eIF1A e eIF3 interagem com a subunidade 40S e o eIF6 com a 60S, mantendo as subunidades dissociadas. O fator eIF2 forma um complexo ternário com o GTP e o metionil-tRNA iniciador, sendo transferido, juntamente com o eIF1, para a subunidade 40S ligada ao eIF3 e eIF1A formando o complexo de pré-iniciação 43S. O fator eIF4F, cuja subunidade eIF4E reconhece especificamente o 7-metil GTP da extremidade 5' do mRNA, atua na interação deste complexo com o mRNA. O códon de iniciação da tradução é selecionado com a participação direta do metionil-tRNA iniciador. O eIF5 interage com o complexo de pré-iniciação e ativa a hidrólise do GTP levando à dissociação dos fatores de iniciação da tradução da subunidade 40S. A subunidade 60S se associa formando o complexo de iniciação de 80S, que em seguida entra na fase de elongação da cadeia polipeptídica. O complexo eIF3 está envolvido em aspectos cruciais da iniciação da tradução, sendo responsável pela dissociação do complexo ribossomal 80S e pela estabilização do complexo ternário (eIF2-GTP-metionil-tRNA iniciador). Além disso, promove a ligação do mRNA à subunidade 40S ribossomal (Figura 1.2).

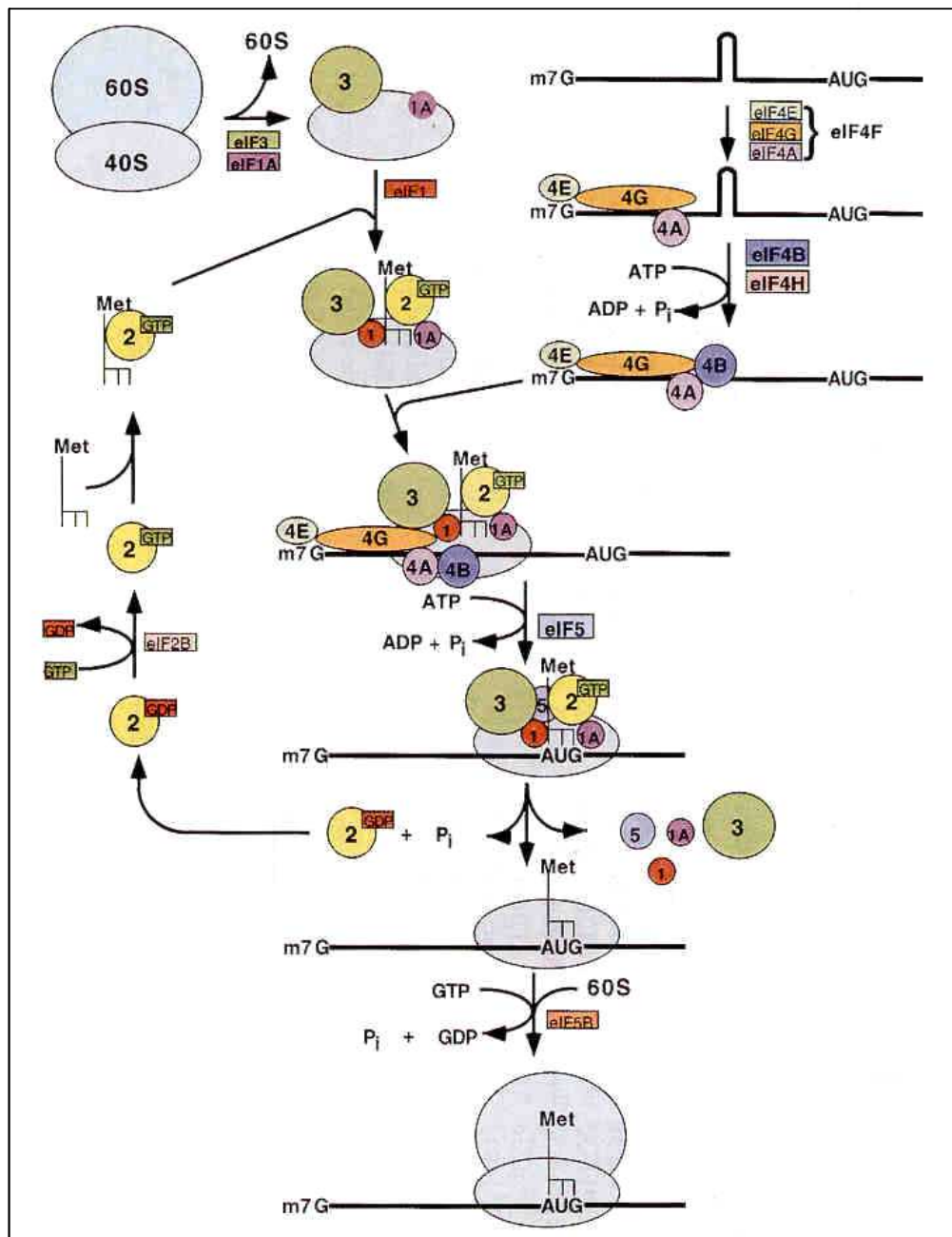


Figura 1.2: Modelo da iniciação da tradução em eucariotos (Hershey & Merrick, 2000). Estão representadas as subunidades 40S e 60S (cinza), assim como os fatores de iniciação da tradução e o mRNA a ser traduzido. 1:eIF1; 1A: eIF1A; 2: eIF2; 3: eIF3; 4A: eIF4A; 4B: eIF4B; 4E: eIF4E; 4G: eIF4G; 5: eIF5; m7G: 7-metil-guanosina.

1.1.2- Localização celular

Estudos sobre a localização celular da INT6 se mostram controversos. Desbois *et al.* (1996) em seu estudo de interação da hINT6 com a proteína Tax observaram a localização nuclear desta proteína em corpos promielocíticos leucêmicos (PML, *promyelocytic leukemia*). Entretanto, Diella *et al.* (1997) relataram a localização citoplasmática da proteína INT6 de camundongo. Um estudo realizado por Guo & Sen (2000) tentou esclarecer as dúvidas acerca da localização celular da proteína INT6. Foi demonstrado que a proteína INT6 apresenta um domínio de localização nuclear bipartido na região central e um domínio de exportação nuclear na região N-terminal (Figura 1.1) sugerindo que esta proteína pode transitar entre o núcleo e citoplasma.

Watkins & Norbury (2004) sugerem que a maior fração da INT6 é nuclear e não co-localiza com o eIF3, reforçando a hipótese de que a INT6 deva apresentar funções biológicas não relacionadas à tradução. Além disso, foi demonstrada também variação da localização subcelular da hINT6 relacionada à progressão do ciclo celular. Em células quiescentes a hINT6 é predominantemente nuclear, mostrando um padrão mais difuso nas células em divisão, sendo mais evidente nas fases G1 e início da fase S onde cerca de 40% da hINT6 parece citoplasmática. Este estudo também mostrou a utilização da técnica de interferência de RNA diminuindo os níveis da proteína na célula, entretanto não descreveu nenhum fenótipo relacionado a esta diminuição.

1.1.3- A conservação da proteína INT6 e estudos funcionais

A proteína INT6 foi altamente conservada ao longo da evolução. Do ponto de vista estrutural, um aspecto interessante é que a INT6 apresenta o domínio PCI (Proteassomo, complexo COP9/signalossomo, fator 3 de iniciação da tradução) na sua região C-terminal (Figura 1.1). Este domínio é formado por aproximadamente

200 aminoácidos, sendo compartilhado por 5 subunidades do eIF3 (eIF3e/INT6, eIF3a/p170, eIF3c/p110, eIFl/p66, eIF3k), por 5 subunidades do proteassomo e por 6 subunidades do complexo COP9/signalossomo (Glickman *et al.*, 1998; Hofmann & Bucher, 1998; Browning *et al.*, 2001; Morris-Desbois *et al.*, 2001; Sheel & Hofmann, 2005).

O proteassomo é formado por uma subunidade regulatória de 19S e uma subunidade proteolítica de 20S, sendo responsável pela degradação de proteínas ubiquitinadas (deMartino & Slaughter, 1999). O complexo COP9/signalossomo (CSN) foi inicialmente descrito em *Arabidopsis thaliana* por sua participação na transdução de sinal durante a germinação fotoinduzida (Karniol *et al.*, 1998; Wei & Deng, 1998). Entretanto este complexo apresenta ortólogos descritos para outros organismos eucarióticos, incluindo os mamíferos, o que sustenta a hipótese de que o CSN é um regulador geral do desenvolvimento (Kim *et al.*, 2001). Este complexo exibe atividade de quinase e está envolvido em regulação da degradação de proteínas (Schwechheimer & Deng, 2001). Alguns estudos sugerem que o PCI possa estabilizar contatos proteína-proteína dentro dos complexos (Kapelari *et al.*, 2000; Tsuge *et al.*, 2001). Embora não existam estudos estruturais até o momento a respeito da proteína humana INT6, predições de estrutura secundária revelam que esta proteína é formada em grande parte por α -hélices.

As espécies de eucariotos *Schizosaccharomyces pombe* (Bandyopadhyay *et al.*, 2000; Crane *et al.*, 2000; Yen & Chang, 2000 ; Akiyoshi *et al.*, 2001), *Arabidopsis thaliana* (Yahalom *et al.*, 2001), *Drosophila melanogaster* (Miyazaki *et al.*, 1999), *Caenorhabditis elegans* (Morris-Desbois *et al.*, 1999), *Xenopus laevis* (Morris-Desbois *et al.*, 1999), *Mus musculus* (Marchetti *et al.*, 1995) e *Homo sapiens* (Desbois *et al.*, 1996) apresentam ortólogos para a INT6 com alta similaridade de sequência de aminoácidos. A proteína de camundongo apresenta sequência de aminoácidos idêntica à proteína humana (Marchetti *et al.*, 1995). É interessante ressaltar que foi identificada uma possível ortóloga para a proteína INT6 na levedura *S. cerevisiae*, denominada Pci8p (Shalev *et al.*, 2001). Esta proteína apresenta um domínio PCI e interage com o complexo eIF3 de levedura *in vivo* e *in vitro*. Além disso, foi caracterizado que o sítio de ligação de Pci8p ao eIF3 se encontra na extremidade N-

terminal da subunidade Prt1p, que contém um motivo de reconhecimento de RNA (RRM, *Recognition RNA Motif*). A proteína hINT6, além de conter um domínio PCI, também se liga ao complexo de levedura eIF3 e à subunidade humana eIF3b/p116 (ortóloga de Prt1p de levedura), através do RRM N-terminal. Em humanos, a proteína INT6 está presente em grandes quantidades em muitos tecidos, incluindo glândulas mamárias, cérebro, coração, fígado, rins, pulmão, ovários, baço e testículos. Além disso, foi detectado mRNA da INT6 em embriões de camundongos a partir do 8º dia de vida intrauterina (Marchetti *et al.*, 1995).

A descoberta de que a levedura de fissão *S. pombe* apresenta uma proteína ortóloga para a INT6 possibilitou o estudo do papel desta proteína na iniciação da tradução neste organismo. A INT6 de *S. pombe* apresenta 43% de identidade de sequência de aminoácidos e 59% de similaridade em relação à proteína de mamíferos (Bandyopadhyay *et al.*, 2000). Apesar das evidências de que a proteína INT6 possa estar associada com a subunidade 40S ribossomal *in vitro*, (Bandyopadhyay *et al.*, 2000) cepas mutantes Δ int6 se mostraram viáveis apresentando somente uma moderada inibição da taxa global da síntese de proteínas (Bandyopadhyay *et al.*, 2000; Crane *et al.*, 2000; Akiyoshi *et al.*, 2001). Outros fenótipos foram observados: cepas mutantes Δ int6 crescem mais devagar que cepas controles em meio mínimo (Crane *et al.*, 2000; Akiyoshie *et al.*, 2001), apresentam defeitos na segregação dos cromossomos por modificação na polimerização e despolimerização dos microtúbulos (Yen & Chang, 2000), formação de tétrades incompletas na meiose, além de hipersensibilidade à cafeína (Bandyopadhyay *et al.*, 2000). Por outro lado, a superexpressão da INT6 neste organismo causa resistência a diversas drogas como, actinomicina D, cicloheximida e cafeína por um aumento na quantidade de RNA de genes regulados pelo fator de transcrição Pap-1 (fator de transcrição tipo AP-1 humano, Crane *et al.*, 2000). Um estudo interessante realizado por Yen *et al.* (2003) demonstrou que cepas de *S. pombe* Δ int6 apresentavam fenótipos semelhantes a mutantes da “tampa” (*lid*) do proteassomo, sendo hipersensíveis à canavalina, um análogo da arginina que aumenta a demanda de atividade do proteassomo devido ao enovelamento incorreto de proteínas. Os mutantes Δ int6 acumularam proteínas ubiquitinadas sugerindo

alterações funcionais do proteassomo. Esta correlação genética é indicativa de uma interação física. De fato, a hINT6 interage com a subunidade Rpn5 (da “tampa”) em ensaios de duplo-híbrido e *pull-down*. Este estudo sugere um novo papel para a INT6 na regulação da degradação de proteínas através do proteassomo 26S.

1.2- Identificação de ligantes protéicos da proteína INT6

Através do método do duplo-híbrido, interações entre a INT6 e outras proteínas estão sendo detectadas auxiliando o entendimento de sua função. A proteína INT6 de *S. pombe* interage com a proteína Moe1 (*Microtubule over extended*), formando um complexo que atua na segregação correta dos cromossomos (Yen & Chang, 2000). Pelo mesmo método identificou-se a interação da hINT6 com a proteína p56, que é uma proteína citoplasmática induzida por interferon (Guo & Sen, 2000), com a proteína Ret finger, que apresenta os domínios *RING finger*, *B-box zinger finger* e *α -helical coiled-coil* e com a subunidade eIF3c/p110 do eIF3 (Morris-Desbois *et al.*, 1999). Além disso, a hINT6 também interage com subunidades do complexo CSN (Hoareau Alves *et al.*, 2002). O papel fisiológico da proteína INT6 neste complexo ainda não está esclarecido.

O estudo da INT6 se torna de grande interesse tendo em vista os vários papéis que esta proteína pode desempenhar na célula uma vez que apresenta um sinal bi-partido de localização nuclear e um sinal de exportação nuclear (Figura 1.1). Além disso, faz parte do eIF3 e interage com a proteína p56, podendo representar um novo mecanismo de regulação da tradução, já que a expressão desta proteína inibe a tradução tanto *in vivo* quanto *in vitro*. A INT6 pode atuar no controle do ciclo celular, além de poder participar do processo de degradação de proteínas ubiquitinadas através do proteassomo. A interação da proteína INT6 com subunidades do complexo CSN também foi descrita.

Na fase inicial deste projeto foi realizada a triagem de proteínas que interagem com a proteína hINT6 e com seu domínio PCI pelo método do duplo-híbrido de levedura. Este método foi descrito por Fields & Song (1989) e, desde então, tem sido

amplamente usado para analisar a interação entre proteínas *in vivo*, seja para testar a interação entre duas proteínas ou para isolar proteínas de bancos genômicos ou de cDNAs. O método do duplo-híbrido baseia-se no fato de que alguns fatores de transcrição específicos são compostos por dois domínios, um responsável por fazer contato com o DNA (domínio de ligação ao DNA) e o outro por ativar a transcrição (domínio de ativação). Estes domínios não precisam fazer parte de um mesmo polipeptídeo para ativar a transcrição, necessitam apenas estar fisicamente próximos, o que pode ser feito através de interação entre proteínas diferentes. Deste modo, o método do duplo-híbrido consiste na expressão de duas proteínas de fusão, uma contendo o domínio de ligação ao DNA (chamada de isca) e a outra o domínio de ativação da transcrição (chamada de presa), numa mesma cepa de *S. cerevisiae* (Figura 1.3).

Neste trabalho foram usadas, inicialmente, como iscas a proteína hINT6 e o domínio PCI em fusão ao domínio de ligação ao DNA da proteína *lexA* de *E. coli*. O rastreamento foi feito com a cepa L40 (Hollenberg *et al.*, 1995) cujo genoma foi modificado para inserir genes marcadores para detectar interação entre proteínas híbridas. Estes marcadores são o gene *HIS3* da própria *S. cerevisiae* e o gene *lacZ* de *E. coli*. Ambos marcadores estão sob controle de promotores mínimos reconhecidos pela proteína *lexA*, conferindo prototrofia à histidina e coloração azul quando o gene *lacZ* é expresso na presença do substrato X-gal (5-bromo-4-cloro- β -D-galactopiranosídeo). A ativação da transcrição dos genes marcadores depende da interação da hINT6 com as proteínas fusionadas ao domínio de ativação do fator de transcrição *GAL4* de *S. cerevisiae*.

Esta triagem apresentou algumas dificuldades devido ao fato de que tanto a proteína hINT6 como o domínio PCI ativaram os genes repórteres. Apesar disso, foi possível identificar interações já descritas para a hINT6, como a proteína Ret finger (Morris-Desbois *et al.*, 1999) e a subunidade eIF3a/p170 do eIF3 (Asano *et al.*, 1997). Um grande número de prováveis novas interações também foi identificado, embora não tenha sido possível testar a interação direta devido à impossibilidade em se expressar a hINT6 em *E. coli* na forma solúvel.

Como alternativa para concluir a etapa do projeto que visava determinar as proteínas que interagem com a hINT6, foi realizado um novo duplo-híbrido usando como isca a forma truncada hINT6T22. Esta forma truncada contém a região do N-terminal até o aminoácido 317 (Figura 1.1). Como resultado deste duplo-híbrido, foram observadas com maior frequência as proteínas PLZF (*Promyelocytic Leukemia Zinc Finger*, 20 clones), HIRIP3 (*HIRA interacting protein 3*, 7 clones) e Histona H3-família 3A (3 clones). Após a realização dos ensaios de ativação dos genes repórteres por estas presas, verificamos que todas foram capazes de ativar os genes *HIS3* e *lacZ* na ausência da isca hINT6T22. Estes resultados somados a dados conhecidos da literatura sobre a proteína PLZF e sobre a histona H3 fusionada ao domínio de ativação *GAL4* nos levaram a conclusão de que estas interações não são específicas.

Em consequência disto, os estudos com a proteína humana INT6 foram interrompidos e mais ênfase foi dada à análise estrutural da proteína INT6 de *A. thaliana*. Além disso, o possível fator de iniciação da transcrição NY-REN-21 foi escolhido como alvo de estudo. Esta proteína foi isolada no duplo-híbrido onde a proteína hINT6 foi usada como isca. Embora a interação com a hINT6 não tenha sido confirmada, esta proteína se mostrou bastante interessante, visto que foi descrita como um antígeno de carcinoma de células renais (Scanlan *et al.*, 1999) e existem poucas informações a respeito de sua estrutura e papel fisiológico. Além disso, esta proteína é uma possível ortóloga do fator de iniciação da transcrição ZPF38 de camundongo com a qual compartilha 82% de identidade de sequência de aminoácidos. Ambas as proteínas apresentam 7 domínios dedos de zinco C_2H_2 de interação com DNA e um domínio de dimerização denominado SCAN (SRE-ZBP, CTfin-51, AW-1, Number 18, Williams *et al.*, 1995) na porção N-terminal.

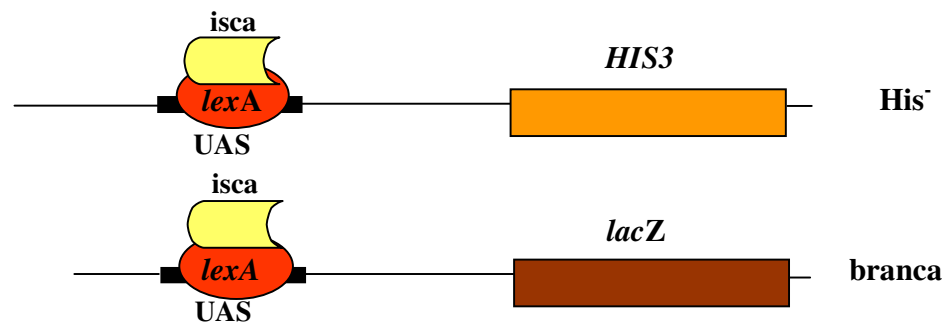
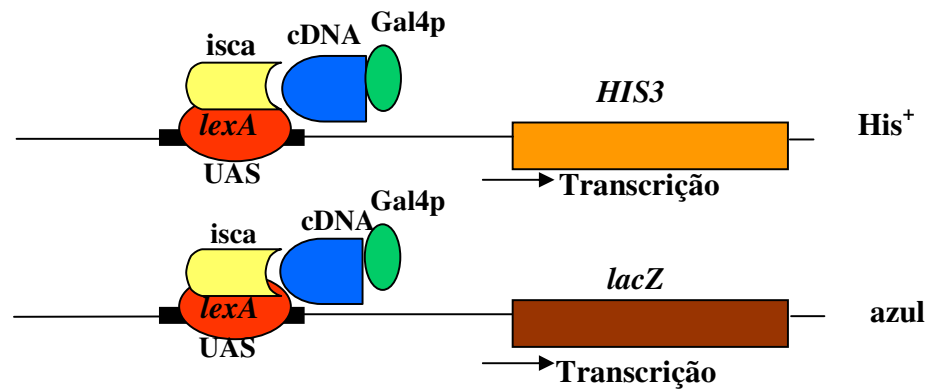
A**B**

Figura 1.3: Diagrama representando o método do duplo-híbrido utilizado para isolar proteínas que interagem com a hINT6, sua forma truncada hINT6T22 e domínio PCI. **A-** Situação em que as duas proteínas de fusão não interagem. Neste caso a proteína *lexA-isca* se liga ao DNA mas não ocorre a transcrição dos genes marcadores *HIS3* e *lacZ*. **B-** Situação em que a porção de proteína de fusão codificada pelo cDNA interage com a hINT6, de maneira que os dois segmentos do fator de transcrição são aproximados, permitindo a transcrição dos genes marcadores pela RNA polimerase II. UAS: Sequência de ativação *upstream*.

1.3- A proteína NY-REN-21

1.3.1- Promotores e maquinaria de transcrição de mRNAs de eucariotos

O núcleo da célula eucariótica contém 3 polimerases diferentes que atuam em funções distintas: RNA polimerase I, transcreve os genes ribossomais (rRNAs), RNA polimerase II, atua na transcrição dos genes que codificam proteínas (mRNAs) e a RNA polimerase III, que transcreve os RNAs transportadores (tRNAs, Zawel & Reinberg, 1995). A transcrição pela RNA polimerase II é um processo que requer a montagem de um complexo de proteínas chamadas fatores basais de iniciação da transcrição na região promotora dos genes (Orphanides *et al.*, 1996; Roeder, 1996; Hampsey, 1998; Dvir *et al.*, 2001).

A região promotora dos genes eucarióticos é formada pelo cerne do promotor com cerca de 35 nucleotídeos adjacentes ao sítio de iniciação da transcrição e pelas seqüências acentuadoras, localizadas mais distantes deste sítio (Roeder, 1991; Tjian & Maniatis, 1994; Smale & Kadonaga, 2003). Embora se acreditasse originalmente que o promotor da RNA polimerase II fosse invariável, foram encontradas diferenças estruturais e funcionais ao longo dos anos. Esta diversidade parece ter importante contribuição para a regulação da expressão gênica (Butler & Kadonaga, 2002).

Os elementos encontrados em um promotor são a seqüência TATA, o elemento iniciador, o elemento promotor *downstream* (DPE, *downstream promoter element*) e a seqüência de reconhecimento do fator de transcrição II B (TFIIB). Um promotor pode ser formado por combinações destas seqüências ou por todas elas (Smale & Kadonaga, 2003). A primeira seqüência identificada em promotor foi a TATA, formada pela seqüência consenso TATAAA. Uma análise em banco de dados de genes humanos, mostrou que esta seqüência está presente em 32% de 1031 possíveis promotores analisados (Suzuki *et al.*, 2001).

Estudos de reconstituição da transcrição *in vitro* usando fatores de transcrição (TF, *Transcription Factor*) purificados revelaram que o processo requer 5 fatores basais: TFIIB, TFIID, TFIIE, TFIIIF, TFIIH e um sexto fator que potencializa a transcrição, TFIIA (Orphanides *et al.*, 1996; Roeder, 1996; Hampsey, 1998). O

primeiro passo para o início da transcrição é o reconhecimento do promotor pelo TFIID, um complexo contendo a proteína ligadora da sequência TATA (TBP, *TATA binding protein*) e pelo menos mais 14 fatores associados (Burley & Roeder, 1996; Albright & Tjian, 2000; Green, 2000). O fator TFIIB estabiliza a ligação de TFIID no promotor por contato com TFIID e seqüências de DNA que flanqueiam a seqüência TATA formando o complexo de pré-iniciação. TFIIB recruta o complexo RNA polimerase II-TFIIF, mas a transcrição só tem início quando TFIIE e TFIIH estiverem incorporados no complexo de pré-iniciação (Figura 1.4). TFIIH tem atividade helicase e separa a fita dupla de DNA no sítio de início da transcrição. Tanto TFIIF e TFIIH são necessários para o escape do promotor e progressão para a fase de alongamento da transcrição (Coin & Egly, 1998; Dvir *et al.*, 2001).

Em adição a este mecanismo de transcrição basal, a regulação da transcrição é governada por proteínas conhecidas como fatores de iniciação da transcrição que estão envolvidos em todos os processos fundamentais biológicos como desenvolvimento, diferenciação, proliferação celular e resposta a estímulos externos. Estes fatores consistem de pelo menos 2 domínios funcionalmente distintos: um domínio de ligação ao DNA, que reconhece e se liga especificamente a seqüências de DNA no promotor de genes específicos e um domínio de ativação ou repressão que influencia a taxa de transcrição pela interação com componentes do complexo de transcrição basal ou outros fatores de transcrição (Ganss & Jheon, 2004, Figura 1.4). Os fatores de transcrição foram classificados de acordo com a natureza de seus domínios de ligação ao DNA, principalmente porque a análise da interação DNA-proteína é tecnicamente menos dispendiosa e mais avançada do que a análise das interações proteína-proteína. Estas classes de fatores de transcrição incluem os motivos de ligação ao DNA como o homeodomínio (Gehring *et al.*, 1994), *paired Box* (Dahl *et al.*, 1997), zíper de leucina (Lamb & McKnight, 1991; Kerppola & Curran 1995; Vison *et al.*, 2002), hélice-alça-hélice (*helix-loop-helix*, Littlewood & Evan, 1995) domínio *runt* (Westendorf & Hiebert, 1999), hélice-volta-hélice (*helix-turn-helix*, Wintjens & Rooman, 1996) e domínio Ets (Oikawa & Yamada, 2003). Entretanto, a maior família de fatores de transcrição contém um domínio de ligação ao DNA conhecido como dedo de zinco, sendo subdividida em várias classes.

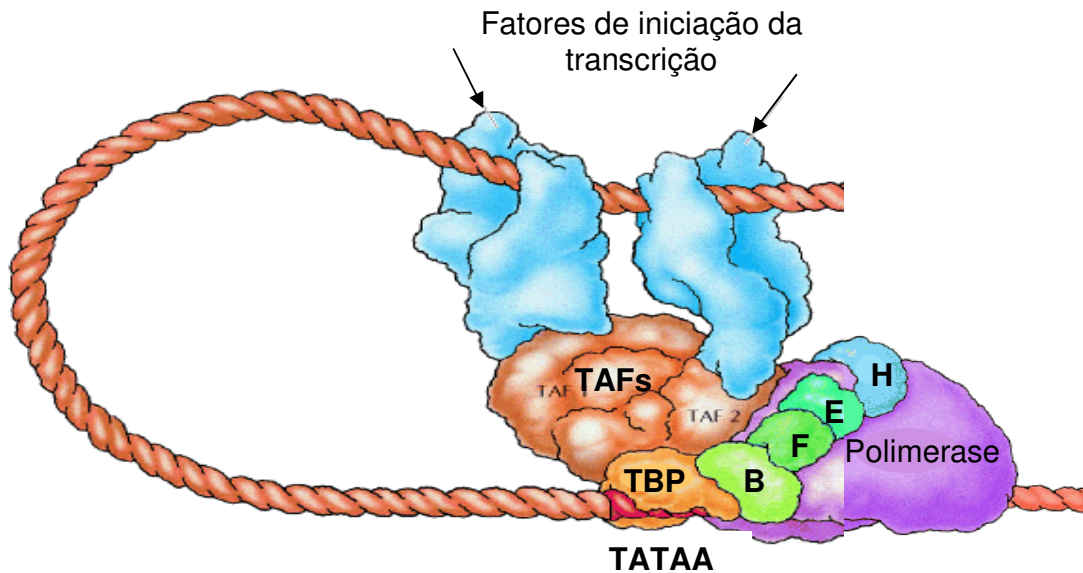


Figura 1.4: Formação do complexo de iniciação da transcrição pela RNA polimerase II (roxo). Muitos promotores apresentam a seqüência consenso TATA. Esta seqüência é reconhecida pela TBP que faz parte do complexo multiprotéico TFIID, representado pela TBP e pelas proteínas associadas a TBP (TAFs). Os fatores TFIIB, F, E e H também se associam ao complexo para o início da transcrição. Fatores de iniciação da transcrição (azul) se ligam a seqüências acentuadoras gene específicas e interagem com os fatores gerais de transcrição ativando ou reprimindo o processo. Fonte: Cooper, 2000.

1.3.2- Proteínas dedos de zinco tipo C₄ e C₂H₂

É estimado que cerca de 3% de todos os produtos de genes humanos contenham 1 ou mais domínios ligadores de zinco (Dawid *et al.*, 1998). Estes domínios são definidos pela presença de uma ou mais moléculas de zinco que estabilizam a estrutura do domínio. Mais de 20 classes diferentes de dedos de zinco foram descritas. Elas diferem no número de moléculas de zinco ligadas e no espaçamento e identidade dos aminoácidos ligantes. O zinco é normalmente coordenado a resíduos de cisteína e histidina, embora o aspartato tenha sido relatado no domínio LIM (Lin-11, Isl-1 e Mec-3, Dawid *et al.*, 1998). A forma mais comum é o clássico C₂H₂, embora a forma C₄ também represente um grupo abundante e importante de fatores de transcrição.

A classe de dedos de zinco tipo C₄ é representada pelos receptores nucleares. O domínio de ligação ao DNA deste receptor é uma das regiões de interação ao DNA mais abundantes. Ele é formado por uma região altamente conservada de 66 aminoácidos localizada na região central dos receptores, juntamente com uma sequência menor e menos conservada que se estende pela região mais flexível da proteína (Rastinejad, 2001). Duas alças que ligam zinco através de 4 cisteínas e um par de α -hélices compõem a estrutura desta região (Luisi *et al.*, 1991, Figura 1.5A).

Os receptores nucleares são uma grande família de fatores de transcrição regulados por ligantes e são importantes na regulação de eventos cruciais do desenvolvimento, diferenciação e homeostase. Membros desta família incluem receptores que respondem a hormônios endócrinos, vitamina D3, hormônios da tireóide e ácido retinóico (Mangelsdorf *et al.*, 1995; Beato & Klung, 2000; Rastinejad, 2001). Outros receptores respondem a intermediários do metabolismo de lipídeos como os receptores ativadores de proliferação de peroxissomos (PPARs) ou a xenobióticos como o receptor X de gravidez (Kliwer *et al.*, 1999; Hihi *et al.*, 2002).

O dedo de zinco C₂H₂ é um pequeno domínio peptídico de ligação ao DNA com estrutura secundária formada por 2 fitas β e 1 α -hélice, estabilizada por um íon de zinco ligado a dois resíduos de cisteína e histidina (Figura 1.5B, Lee *et al.*, 1989). Este domínio interage em torno da dupla hélice do DNA se ligando pela cavidade

maior. O DNA se liga na proteína pelos aminoácidos das posições -1, +3, e +6 (onde +1 é o primeiro resíduo da região de α -hélice, Pavletich & Pabo, 1993).

Os dedos de zinco estão conectados por seqüências de aminoácidos conservados. Aproximadamente metade das proteínas C_2H_2 conhecidas contém a seqüência TGEKP que conecta os dedos de zinco adjacentes (Wolfe *et al.*, 2000; Laity *et al.*, 2000a; Laity *et al.*, 2001). Estudos de mutagênese sugeriram que estas seqüências apresentam um papel importante na ligação ao DNA, sendo que mutações únicas podem reduzir a afinidade de ligação em mais de 20 vezes (Choo & Klug, 1993). Estas seqüências se apresentam móveis e não estruturadas na ausência da cadeia específica de DNA. Após a ligação da seqüência correta de DNA, os aminoácidos de ligação perdem a mobilidade e se estruturam em um *cap* de hélice C-terminal (Laity *et al.*, 2000b).

As proteínas dedos de zinco podem se ligar ao DNA e regular a expressão dos genes alvos de forma positiva ou negativa (Laity *et al.*, 2001). Além da interação com ácidos nucleicos, interações entre proteínas mediadas pelos dedos de zinco também foram descritas (Sun *et al.*, 1996; Morgan *et al.*, 1997; McCarty *et al.*, 2003).

A diversidade dentro desta família de genes é encontrada em domínios modulares conservados que estão fora da região dos dedos de zinco. Estes domínios geralmente controlam a associação seletiva dos fatores de transcrição com outros fatores celulares. Estas interações são freqüentemente essenciais para a regulação, localização subcelular, ligação ao DNA ou transcrição. Os domínios conservados encontrados no N-terminal destes fatores de transcrição incluem o BTB/POZ (*Broad-Complex*, *Tramtrack*, *Bric-a-brac/Poxvirus* and *Zinc finger*, Bardwell & Treisman, 1994), KRAB (*Krüppel-Associated Box*, Bellefroid *et al.*, 1991), ZAD (*Zinc Finger Associated C4DM*, Chung *et al.*, 2002) e SCAN (*SRE-ZBP*, *CTfin-51*, *AW-1*, *Number 18*, Williams *et al.*, 1995), também chamado LeR (*Leucine Rich*, Pengue *et al.*, 1994). Este último é encontrado na proteína ZFP38/NY-REN-21.

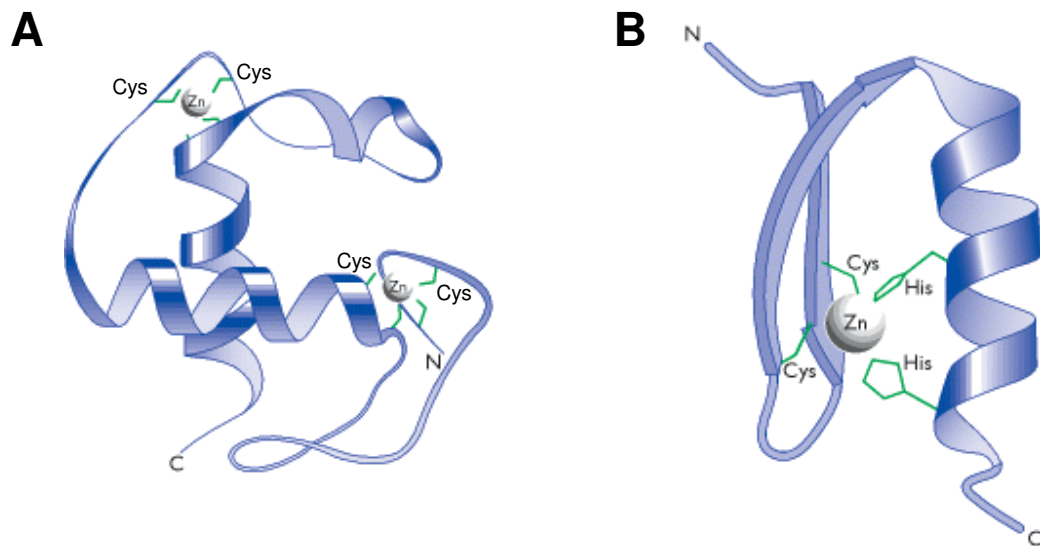


Figura 1.5: Estrutura de regiões de interação ao DNA tipo dedo de zinco. **A-** Receptor de hormônio esteróide (dedo de zinco tipo C_4). **B-** Proteína de levedura SWI5 (dedo de zinco tipo C_2H_2). N e C: regiões amino e carboxi-terminais da proteína, respectivamente. Cys: cisteína, His: histidina, Zn: íon de zinco. O grupamento R dos aminoácidos envolvidos na interação com o átomo de zinco estão representados por linhas verdes. Fonte: Brown, 2002.

1.3.3- Domínios associados às proteínas dedos de zinco

1.3.3.1- Domínio KRAB

O domínio KRAB é encontrado na porção N-terminal de centenas de proteínas dedos de zinco tipo C_2H_2 (Bellefroid *et al.*, 1991; Rosati *et al.*, 1991) e parece estar restrito aos vertebrados (Looman *et al.*, 2002). Tem sido relatado como um potente repressor da transcrição (Margolin *et al.*, 1994; Witzgall *et al.*, 1994; Senatore *et al.*, 1999). Quando em fusão a motivos de ligação ao DNA, este domínio reprime tanto a transcrição basal quanto a atividade de transcrição em células transientemente transfectadas com expressão dose dependente (Pengue *et al.*, 1994; Deuschle *et al.*, 1995; Moosmann *et al.*, 1997).

A atividade repressora do domínio KRAB é mediada pelo co-repressor TIF1 β , também chamado KAP-1 ou Krip-1 (Friedman *et al.*, 1996; Moosmann *et al.*, 1996).

Este co-repressor pertence ao grupo de proteínas que possuem o domínio RBCC (*Ring Finger*, *B-box*, *Coiled coil*), sendo recrutado pelas proteínas contendo domínio KRAB através de interação direta proteína-proteína entre os domínios KRAB e RBCC (Peng *et al.*, 2000a; Peng *et al.*, 2000b).

O domínio KRAB consiste de aproximadamente 75 aminoácidos distribuídos em 2 motivos adjacentes denominados A (aproximadamente 42 aminoácidos) e B (aproximadamente 33 aminoácidos), que são codificados por exons distintos submetidos a *splicing* diferencial (Rosati *et al.*, 1999; Rosati *et al.*, 2001). O módulo A apresenta atividade de repressão da transcrição (Margolin *et al.*, 1994; Witzgall *et al.*, 1994; Senatore *et al.*, 1999), enquanto o módulo B parece estar restrito ao aumento de atividade do módulo A (Vissing *et al.*, 1995; Poncelet *et al.*, 1998).

Curiosamente, análises por gel filtração, dicroísmo circular e espectrometria de infravermelho evidenciaram que tanto o módulo KRAB-A, quanto o módulo KRAB-AB apresentam estrutura pouco compacta, desordenada, flexibilidade e tendência a agregação (Mannini *et al.*, 2006). Juntamente com importantes papéis regulatórios na célula como diferenciação e desenvolvimento (Collins *et al.*, 2001), as proteínas dedos de zinco contendo o domínio KRAB parecem estar envolvidas na inibição do potencial oncogênico de algumas proteínas como o supressor de tumor Von Hippel-Lindau (Li *et al.*, 2003) e c-myc (Hennemann *et al.*, 2003).

1.3.3.2- Domínio BTB/POZ

O domínio BTB também chamado POZ foi originalmente identificado como um domínio conservado presente em complexos reguladores da transcrição em *D. melanogaster* e em muitas proteínas dedos de zinco do vírus pox (Koonin *et al.*, 1992; Bardwell & Treisman, 1994; Zollman *et al.*, 1994). Vários papéis têm sido atribuídos a este domínio, incluindo repressão da transcrição (Melnick *et al.*, 2000; Ahmad *et al.*, 2003), regulação do citoesqueleto (Bomont *et al.*, 2000; Ziegelbauer *et al.*, 2001) e ubiquitinação e degradação de proteínas (Pintard *et al.*, 2003; Xu *et al.*, 2003; Kobayashi *et al.*, 2004; Wilkins *et al.*, 2004).

Na espécie humana grande parte das proteínas que apresentam domínio BTB/POZ também contém dedos de zinco na porção C-terminal (Stogios *et al.*, 2005). Estas duas regiões são ligadas por uma longa região conectora que varia de 100 a 375 aminoácidos. Esta região é a que apresenta menor conservação de seqüência primária da proteína sendo os dedos de zinco mais conservados seguidos do domínio BTB. A região conectora apresenta baixa complexidade de seqüência e é predita como desestruturada na maioria dos casos (Stogios *et al.*, 2005). Muitos membros desta grande família foram caracterizados como importantes fatores transcricionais e alguns estão envolvidos no desenvolvimento celular maligno como BCL6 (*B-cell lymphoma* 6, Polo *et al.*, 2003; Albagli-Curiel, 2003) LRF (*lymphoma related factor*, Maeda *et al.*, 2005), PLZF (Costoya & Pandolfi, 2001), HIC (*hypermethylated in cancer 1*, Chen *et al.*, 2004; Pinte *et al.*, 2004) e MIZ1 (*Myc interacting zinc finger 1*, Peukert *et al.*, 1997). Associações heterodiméricas entre as proteínas contendo BTB foram descritas (Hoatlin *et al.*, 1999; Takenaga *et al.*, 2003). A formação de heterodímeros nos fatores BTB pode ser um mecanismo de direcionar estas proteínas a elementos regulatórios particulares pela combinação de diferentes associações de cadeias de domínios ligadores de DNA gerando especificidades distintas de reconhecimento de DNA (Kobayashi *et al.*, 2000).

1.3.3.3- Domínio ZAD

O domínio ZAD está restrito às proteínas dedos de zinco de insetos, sendo formado por cerca de 71 a 97 aminoácidos divididos em 4 blocos de seqüências separadas por 3 repetições variáveis de tamanhos diferentes. A função deste domínio ainda não está descrita, mas acredita-se que pode mediar interações entre proteínas e que proteínas contendo este domínio podem atuar no controle da transcrição (Chung *et al.*, 2002).

1.3.3.4- Domínio SCAN

Foram descritos 71 genes contendo o domínio SCAN no genoma humano. Este grupo de genes representa 10% dos 700 genes contendo o domínio de dedos de zinco C_2H_2 descritos no genoma humano (Sander *et al.*, 2003). O domínio SCAN está restrito aos vertebrados (Lander *et al.*, 2001; Venter *et al.*, 2001; Sander *et al.*, 2003) e associado à função de dimerização entre as proteínas que o contém, podendo gerar homodímeros ou heterodímeros (Williams *et al.*, 1999; Sander *et al.*, 2000; Schumacher *et al.*, 2000; Stone *et al.*, 2002).

O domínio SCAN foi descrito por muitos autores (Pengue *et al.*, 1994; Sander *et al.*, 2000; Collins *et al.*, 2001; Honer *et al.*, 2001; Stone *et al.*, 2002), após múltiplos alinhamentos de sequência, como sendo um domínio protéico altamente conservado e formado por cerca de 82 a 88 aminoácidos. A conservação da sequência primária revelou, entretanto, uma divergência entre a região amino-terminal bem conservada e a região menos conservada no C-terminal. Nam *et al.* (2004) propuseram o tamanho mínimo funcional deste domínio de 58 aminoácidos, baseados em algumas observações: 1- Os genes Leap1/PGC2 de camundongos e SDP1/RAZ1/SCAND1 de humanos caracterizam proteínas com a região carboxi-terminal do domínio SCAN menor, gerando um domínio de tamanho final de 60 aminoácidos (Castillo *et al.*, 1999; Sander *et al.*, 2000; Schumacher *et al.*, 2000); 2- A caracterização modular dos domínios SCAN preditos das proteínas ZNF94 e ZNF99 se sobrepõem à região C-terminal com os domínios KRAB adjacentes; 3- O tamanho mínimo proposto por estes autores deve se estruturar de forma adequada e detém as propriedades bioquímicas de dimerização, como previamente foi descrito para os domínios SCAN de maior tamanho.

Estruturalmente, o domínio SCAN representa um módulo de proteína independentemente estruturado com 5 α -hélices distintas em cada monômero. As duas maiores hélices (H1 e H2) formam a interface do dímero, sendo a hélice H1 estruturada contra a hélice H2 do outro monômero e vice-versa. Apesar de não haver similaridade de sequência primária, o dímero SCAN adota uma estrutura similar ao domínio C-terminal da proteína do capsídeo retroviral do vírus da imunodeficiência

humana tipo-1 (HIV-1), embora a interface de dimerização seja bastante diferente (Ivanov *et al.*, 2005).

1.3.4- Identificação e aspectos da proteína NY-REN-21

A proteína NY-REN-21 foi identificada como um antígeno tumoral através de anticorpos autólogos em pacientes com carcinoma de células renais pelo método referido como SEREX (*serological analysis of recombinat cDNA expression libraries*, Scanlan *et al.*, 1999). A função desta proteína ainda não está caracterizada, entretanto representa uma possível ortóloga para a proteína ZFP38 de camundongo, também chamada de CTfin51, Zipro1 e Ru49, com 82% de identidade de sequência de aminoácidos. Em camundongos, a proteína ZFP38 contém 3 seqüências repetidas *in tandem* que funcionam como ativadoras da transcrição quando fusionadas ao domínio GAL4 de ligação ao DNA sendo, portanto, uma proteína regulatória (Chowdhury *et al.*, 1992). Esta proteína se liga ao DNA na seqüência consenso 5' AGTAC 3', sendo a interação mais forte em sítios contendo 2 seqüências com espaçamento preferencial de 2 ou 7 nucleotídeos (Yang *et al.*, 1996). Uma importante diferença observada entre as duas possíveis ortólogas é que a proteína NY-REN-21 não apresenta 77 aminoácidos na região N-terminal, descrita como a responsável pela ativação da transcrição na proteína ZFP38. Análises de *Northern blot* de 16 tecidos revelaram a expressão do transcrito da proteína NY-REN-21 em quase todos, com exceção do timo e maior nível nos testículos (Scanlan *et al.*, 1999).

Estudos funcionais da proteína ZFP38 em camundongos revelaram que este fator de transcrição é um marcador para a linhagem neuronal cerebelar granular. A proteína é expressa desde as primeiras células granulares progenitoras nos lábios rômnicos até a diferenciação e maturação dos neurônios granulares (Yang *et al.*, 1996). Um estudo com camundongos transgênicos que apresentavam aumento do número de cópias do gene e dos níveis de expressão desta proteína mostraram que a ZFP38 tem um papel na proliferação de precursores das células granulares do

desenvolvimento do cerebelo. O aumento de dosagem deste gene também resulta em perda de pêlo associado com aumento da proliferação de células do epitélio e desenvolvimento anormal do folículo capilar (Yang *et al.*, 1999).

Do ponto de vista estrutural, tanto a proteína humana quanto a de camundongo apresentam 7 possíveis domínios dedos de zinco C_2H_2 de interação com DNA, na região C-terminal e o domínio SCAN na porção N-terminal (Figura 1.6). As regiões centrais destas proteínas apresentam cerca de 150 aminoácidos sem estrutura secundária ordenada de acordo com previsões realizadas em programas computacionais como PSIPRED (Jones *et al.*, 1999; McGuffin *et al.*, 2000, <http://bioinf.cs.ucl.ac.uk/psipred>) e PONDR (Romero *et al.*, 2001, www.pondr.com). Além destes domínios, estas proteínas também apresentam uma região acídica na porção N-terminal responsável pela ativação da transcrição na proteína ortóloga de camundongo (Chowdhury *et al.*, 1992, Figura 1.6).

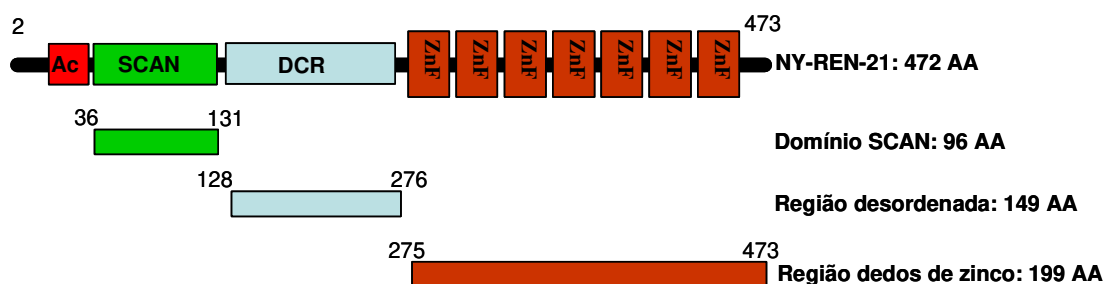


Figura 1.6: Representação da proteína NY-REN-21, seus domínios e regiões clonados para expressão, purificação e caracterização. Os números representam os aminoácidos que flanqueiam o início e o fim das regiões em estudo. Ac: região acídica; SCAN: SRE-ZBP, CTfin 51, Aw-1 e Number 18; DCR: Região desordenada; ZnF: dedos de zinco.

2- OBJETIVOS

Este trabalho teve como meta estudar os aspectos estruturais das proteínas NY-REN-21 e INT6, além da identificação de seus possíveis ligantes protéicos. Desta forma, os seguintes objetivos foram estabelecidos:

- 1) Identificar as proteínas que interagem com a proteína humana INT6 através do método do duplo-híbrido da levedura.
- 2) Caracterização estrutural da proteína INT6 de *A. thaliana* por dicroísmo circular.
- 3) Expressar e purificar a proteína NY-REN-21, assim como seus domínios para análises estruturais.
- 4) Determinar a capacidade de oligomerização do domínio SCAN da proteína NY-REN-21.
- 5) Esclarecer aspectos funcionais da proteína NY-REN-21 através da identificação, pelo método do duplo-híbrido de leveduras, das proteínas com as quais possa formar heterodímeros.

3- RESULTADOS

Os resultados deste trabalho estão apresentados na forma de um capítulo descritivo do duplo-híbrido realizado para a identificação dos ligantes protéicos da proteína INT6 humana, além de 2 artigos. Os artigos representam os resultados das análises estruturais da proteína INT6 ortóloga de *A. thaliana* e a caracterização estrutural e os estudos das interações da proteína NY-REN-21. Mais dois artigos são apresentados como anexos, uma vez que representam o resultado de estudos das proteínas Rab11b e alfa4 isoladas através do duplo-híbrido de levedura usando a proteína INT6 humana como isca.

3.1- Resultados do rastreamento dos ligantes protéicos da hINT6

Identificação dos ligantes protéicos da proteína humana INT6 pelo método do duplo de levedura

Flávia R. G. Carneiro, Thais Haline-Vaz, Tereza C. Lima Silva e Nilson I. T. Zanchin

Identificação dos ligantes protéicos da proteína humana INT6 pelo método do duplo de levedura

Flávia R. G. Carneiro^{‡¶}, Thais Haline Vaz^{‡¶}, Tereza C. Lima Silva[‡] e Nilson I. T. Zanchin^{‡¶}

[‡]Centro de Biologia Molecular Estrutural- Laboratório Nacional de Luz Sincrotron-LNLS

[¶]Departamento de Genética e Evolução-Universidade Estadual de Campinas-UNICAMP

RESUMO

A proteína INT6 foi identificada como a subunidade de 48 kDa do fator 3 eucariótico de iniciação da tradução e triada pelo método do duplo híbrido de levedura como ligante protéico da oncoproteína Tax do vírus tipo I da leucemia humana de células T (HTLV-1). Além disso, o gene que codifica a proteína INT6 corresponde a um dos sítios de inserção do vírus de tumor de glândula mamária em camundongo (MMTV- *Mouse Mammary Tumor Virus*). A inserção do genoma viral pode levar à expressão de proteínas truncadas sem a porção C-terminal. Nestas proteínas o domínio PCI, descrito como domínio de interação entre proteínas, está ausente. Neste trabalho foram realizadas três triagens pelo método do duplo-híbrido de levedura para a identificação dos ligantes protéicos da proteína humana INT6 (hINT6). Foram usados como iscas o cDNA da proteína INT6, de seu domínio PCI e da forma truncada T22 (hINT6T22) em fusão ao domínio de ligação ao DNA *lexA*. Foi isolada uma ampla gama de proteínas distintas com funções diferentes nas triagens usando a proteína hINT6 e o domínio PCI como iscas. Foram realizados testes de interação *in vitro* com proteínas recombinantes, mas não foi possível a confirmação destas interações. Tanto a proteína hINT6 quanto o domínio PCI ativaram os genes repórteres do sistema de duplo-híbrido o que pode ter dificultado bastante o isolamento de interações específicas. A proteína hINT6T22 não auto-ativou os genes repórteres, entretanto o duplo-híbrido usando esta isca só identificou proteínas falso-positivas, sugerindo que esta forma truncada não enovelou corretamente ou não foi translocada para o núcleo ou não é funcional, perdendo a capacidade de interação com outras proteínas, visto que esta forma truncada não apresenta o domínio PCI.

Palavras chaves: INT6; Duplo-híbrido de levedura; domínio PCI.

INTRODUÇÃO

Os genes INT foram identificados como *loci* chaves na formação do câncer de mama devido ao fato de serem freqüentes sítios de inserção do vírus de tumor de glândula mamária em camundongos (MMTV, *Mouse Mammary Tumor Vírus*; revisado por Tekmal & Keshava, 1997). Até o momento, 9 genes INT foram descritos. Entre estes genes estão famílias importantes de moléculas sinalizadoras bastante conservadas como Wnt (codificada pelos genes WNT/INT1, WNT3/INT4 e WNT10b), Notch (codificada pelo INT3), FGF (codificada por FGF3/INT2, FGF4 e FGF8), e a proteína INT6, alvo de nosso estudo (Nusse & Varmus, 1982; Dickson *et al.*, 1984; Gallahan & Callanhan, 1987; Peters *et al.*, 1989; Marchetti *et al.*, 1995).

O cDNA que codifica a proteína humana INT6 também foi isolado em uma triagem através do método do duplo-híbrido de levedura usando a oncoproteína Tax do vírus tipo I da leucemia humana de células T (HTLV-1) como isca (Desbois *et al.*, 1996). Além disso, a proteína INT6 foi caracterizada como a subunidade p48 do fator 3 eucariótico de iniciação da tradução (eIF3, Asano *et al.*, 1997).

A integração do genoma do MMTV no gene INT6 foi descrita nos introns 5 ou 9 na orientação transcricional oposta a da proteína INT6, o que pode levar a formação dos tumores T1139 ou T22, respectivamente. A transcrição dos alelos afetados é terminada precocemente por um sinal de terminação inserido pelo genoma viral podendo levar à expressão de proteínas truncadas sem a porção C-terminal (Marchetti *et al.*, 1995).

A superexpressão da forma truncada hINT6T1139 em células epiteliais e fibroblastos de mamíferos conferiu a estas células a capacidade de crescimento independente de adesão mesmo na presença de 2 alelos funcionais do gene, sugerindo que esta proteína truncada atue de maneira dominante (Rasmussen *et al.*, 2001; Mayeur & Hershey, 2002). Entretanto, Miyazaki *et al.* (1997) mostraram que a INT6 é freqüentemente perdida em carcinomas primários de mama

(30% de todos os pacientes examinados), sugerindo que a perda de função da INT6 contribui para a transformação maligna.

A INT6 faz parte do complexo eIF3, entretanto análises de cepas mutantes Δ int6 em *Schizosaccharomyces pombe* revelaram que a INT6 não é uma subunidade essencial do eIF3 neste organismo (Bandyopadhyay *et al.*, 2000; Akiyoshi *et al.*, 2001), embora seja necessária para a associação estável das subunidades deste complexo (Bandyopadhyay *et al.*, 2002). Estudos em *S. pombe* também revelaram que a proteína INT6 é muito conservada, visto que a proteína humana INT6, mas não suas formas truncadas, foi capaz de recuperar os defeitos de crescimento observados em cepas Δ int6 (Yen & Chang, 2000).

Diversos ligantes protéicos da INT6 foram identificados, entre eles estão as proteínas Ret finger (Morris-Desbois *et al.*, 1999), Moe1 (*Microtubule over extended*, Yen & Chang, 2000) e p56 (Guo & Sen, 2000). A interação com a p56 pode representar um novo mecanismo de regulação da tradução, já que a expressão desta proteína inibe a tradução tanto *in vivo* quanto *in vitro*, o que sugere que a INT6 possa desempenhar uma atividade regulatória do fator de tradução eIF3 (Guo *et al.*, 2000). Um aspecto interessante é que a proteína INT6 também interage com a subunidade Rpn5 do proteassomo em ensaios de duplo-híbrido e *pull down* (Yen *et al.*, 2003), o que sugere novo papel para a INT6 na regulação e degradação de proteínas ubiquitinadas.

Um sinal bipartido de localização nuclear e um sinal de exportação nuclear na porção N-terminal da INT6 foram descritos, sugerindo que esta proteína pode transitar entre o núcleo e o citoplasma da célula (Guo & Sen, 2000). A proteína INT6 também apresenta o domínio conservado PCI (Proteassomo, complexo COP9-signalossomo, fator 3 de iniciação da tradução) na sua região C-terminal. Este domínio é formado por aproximadamente 200 aminoácidos, sendo compartilhado por 5 subunidades do eIF3 (eIF3e/INT6, eIF3a/p170, eIF3c/p110, eIF1/p66, eIF3k), por 5 subunidades do proteassomo e por 6 subunidades do complexo COP9/signalossomo (Hofmann & Bucher, 1998; Glickman *et al.*,

1998; Browning *et al.*, 2001; Morris-Desbois *et al.*, 2001; Sheel & Hofmann, 2005).

Os dados publicados levantaram várias hipóteses sobre a função da proteína INT6. Esta proteína pode estar envolvida em processos de proliferação celular e transformação maligna das células, desempenhar um papel no núcleo em associação com a proteína Ret finger, participar da tradução como parte de eIF3 e do processo de degradação de proteínas ubiquitinadas através do proteassomo.

Neste trabalho, foram realizadas triagens através do método do duplo-híbrido de levedura, com o objetivo de identificar ligantes protéicos da INT6 e contribuir para o esclarecimento do papel da INT6 na célula. Foram usadas como iscas a proteína hINT6, o domínio PCI e a forma truncada hINT6T22. Os rastreamentos com a proteína hINT6 e o domínio PCI revelaram uma ampla gama de proteínas com funções diferentes. Curiosamente, o duplo-híbrido realizado com a forma hINT6T22 isolou apenas proteínas falso-positivas, o que sugere que esta forma da proteína INT6 não enovelou corretamente ou não foi translocada para o núcleo ou que a porção deletada, domínio PCI, seja imprescindível para a interação entre proteínas.

MATERIAL E MÉTODOS

Construção dos vetores

Para a construção dos vetores pBTM-INT6 e pBTM-PCI, o cDNA da proteína humana INT6, cujo código de acesso no GenBank é AI 879234, e do domínio PCI (nucleotídeo 865 até o códon de terminação do cDNA da proteína INT6) foram amplificados por PCR. Os sítios de restrição para as enzimas *EcoRI* e *BamHI* foram adicionados nas extremidades 5' e 3' das seqüências dos cDNAs através dos pares de oligonucleotídeos ONZ31 e ONZ3, para a proteína INT6, e OZN 31 e OZN33, para o domínio PCI (Tabela 1). Os produtos de PCR foram inseridos nos sítios de restrição *EcoRI* e *BamHI* do vetor pBTM116 (Bartel *et al.*, 1993).

A forma truncada hINT6T22 da proteína INT6 (nucleotídeos 4 a 973) foi clonada no vetor pTL-1. Este vetor é derivado do vetor pBTM116,

apresentando a troca da marca de seleção do antibiótico ampicilina para canamicina (Carneiro *et al.*, 2006). Um fragmento *EcoRI*-*BamHI*, amplificado por PCR usando o vetor pBTM-INT6 como *template* e os oligonucleotídeos ONZ31 e ONZ187 (Tabela 1), foi inserido no vetor pTL-1, gerando o vetor pTL-hINT6T22.

O inserto para a construção do vetor pGEX-INT6 foi obtido a partir de digestão do vetor pBTM-INT6 com as enzimas de restrição *EcoRI* e *SalI*. O vetor pGEX-4T-1 (Amersham-Biosciences) foi previamente digerido com as enzimas citadas e o inserto inserido nestes sítios de restrição. Para a construção do vetor pET-protA-INT6, o cDNA que codifica a fusão proteínaA-INT6 foi isolado do vetor de levedura YCp33-protA-INT6 por digestão com as enzimas de restrição *NdeI* e *HindIII* e clonado nos mesmos sítios de restrição do vetor de expressão em *E. coli* pET-23a (Novagen). O vetor YCp33-protA-INT6 foi construído pela amplificação de um fragmento *EcoRI*-*HindIII* que codifica a proteína INT6 pelos oligonucleotídeos ONZ87 e ONZ88 (Tabela 1). Para esta clonagem foi usado o vetor YCp33-GAL-ProtA-RRP43 (Zanchin & Goldfarb, 1999) do qual a porção codificadora da proteína Rrp43p foi retirada por clivagem com as enzimas de restrição *HindIII* e *EcoRI*. A construção do vetor YCp33protA-INT6 foi realizada em dois passos: o primeiro foi a retirada da seqüência codificadora da proteína Rrp43p e inserção do fragmento de interesse nos sítios de *EcoRI* e *HindIII* do vetor YCplac33 (Gietz & Sugino, 1988) e o segundo foi uma nova inserção do fragmento correspondente ao promotor *GAL1* e à proteína A. Este fragmento foi inserido novamente no vetor contendo em suas extremidades sítios de inserção para a enzima *EcoRI*.

Para a construção dos vetores contendo as proteínas 14-3-3 τ , alfa4, IkBL e NY-REN-21 em fusão com cauda de histidina, os cDNAs completos destas proteínas foram amplificados a partir da biblioteca de cDNA de cérebro fetal humano da Clontech (Cat n° HL4028AH). Para a amplificação do cDNA da proteína 14-3-3 τ (código de acesso no GenBank: NM_006826) foram usados os oligonucleotídeos ONZ105 e ONZ106 (Tabela 1), e para a proteína alfa4 (código de acesso no GenBank: NM_001551) os oligonucleotídeos ONZ107 e ONZ108 (Tabela 1). Os cDNAs resultantes contêm os sítios de restrição *NdeI* e *BamHI* nas extremidades 5' e 3', respectivamente. O cDNA da proteína IkBL (código de acesso no GenBank: XM_041842) foi amplificado com a utilização dos oligonucleotídeos ONZ109 e ONZ110 (Tabela 1), inserindo os sítios de restrição

para as enzimas *NdeI* e *EcoRI* nas extremidades 5' e 3' desta sequência, respectivamente. O cDNA do antígeno NY-REN-21 (código de acesso no GenBank: NM_145914) foi amplificado com oligonucleotídeos OZN111 e OZN112 (Tabela 1) gerando um fragmento cujas extremidades 5' e 3' contêm os sítios de restrição das enzimas *NdeI* e

EcoRI. Os insertos amplificados por PCR, assim como o vetor pET-28a (Novagen), foram digeridos com as enzimas adequadas e, posteriormente, os produtos de PCR foram ligados neste vetor. A sequência de DNA foi conferida após sequenciamento dos vetores.

Tabela 1: Lista dos oligonucleotídeos usados nas reações de PCR. Em negrito estão destacados os sítios para as enzimas de restrição.

oligonucleotídeo	sequência	sítio de restrição
OZN31	5' TGG CGAATTC GC GCGAGTACGACTTGACTAC 3'	<i>EcoRI</i>
ONZ3	5' TTAT GGATCCT CAGTAGAAGCCAGA 3'	<i>BamHI</i>
ONZ33	5' TTAC GAATTC AAAGACCCAATTACAGAATTTG 3'	<i>EcoRI</i>
ONZ187	5' TTC AGGATCCT CATGATTACATTCCTCAGC 3'	<i>BamHI</i>
ONZ87	5' GG CGAATTC CGCGGAGTACGACTTGACTAC 3'	<i>EcoRI</i>
ONZ88	5' TATGA AAGCTT CAGTAGAAGCCAGAATC 3'	<i>HindIII</i>
ONZ105	5' G CCCCATATG GAGAAAGACTGAGCTGATCCA 3'	<i>NdeI</i>
ONZ106	5' A AGGATCCC ACCCTGTATGGATTTAGTTTT 3'	<i>BamHI</i>
ONZ107	5' C CCCCATATG GCTGCTGAGGACGAGTTAC 3'	<i>NdeI</i>
ONZ108	5' TGG GATCCT CAGCCCATGTTCTGTCTCGGTTTC 3'	<i>BamHI</i>
ONZ109	5' AGGT CATATG AGTAACCCCTCCCCCAGG 3'	<i>NdeI</i>
ONZ110	5' CGAATTC CCTAGGGTCACTTGAGGGCCTCTG 3'	<i>EcoRI</i>
ONZ111	5' TTAC CATATG ACCAAGGTACTAGGCATGGC 3'	<i>NdeI</i>
ONZ112	5' TGA ATTC TTACGGTGCTTCTCCCTCTCCAG 3'	<i>EcoRI</i>

Duplo-híbrido de levedura

Para a identificação de proteínas que interagem com a INT6, foi utilizado o método do duplo-híbrido de levedura (Fields & Song, 1989) e uma biblioteca de cDNA de cérebro fetal humano clonada no vetor pACT2 (Clontech, Cat. N° HL4028AH). Inicialmente foi realizado o pré-inóculo da cepa L40 (*MATa his3Δ200 trp1-901 leu2-3,112 ade2 lys2-801am LYS2::(lexAop)4-HIS3 URA3::(lexAop)8-lacZ*, Hollenberg *et al.*, 1995) transformada com os vetores pBTM-INT6, pBTM-PCI e pTL-hINT6T22 por 16 horas em 400 mL de meio mínimo YNB (*Yeast Nitrogen Base*) suplementado com adenina, histidina e leucina. Trezentos e trinta mililitros deste pré-inóculo, cujo valor da densidade óptica medida a 600 nm (OD_{600}) estava em 0,9, foram transferidos para um total de

1,5 L de meio YPD (*Yeast extract, Peptone, Dextrose*) suplementado com adenina. A cultura foi, então, incubada a 30°C sob agitação de 225 rpm. Ao atingir a OD_{600} de 0,5, a cultura foi centrifugada a 5000 g por 5 minutos à temperatura ambiente e as células foram ressuspensas em 500 mL da água milliQ estéril. Nova centrifugação foi realizada nas mesmas condições. O sobrenadante foi descartado e as células ressuspensas em 8 mL de uma solução contendo 0,1 M de acetato de lítio e TE 1X (10 mM Tris pH 7,5, 1 mM de EDTA). A este volume de células foram adicionados 0,1 mg da biblioteca, 20 mg de DNA carreador, com posterior agitação em aparelho tipo vórtex e 60 mL de uma solução 0,1 M de acetato de lítio pH 7,0, TE 1X e 40% polietilenoglicol (PEG) 3350. As células foram incubadas a 30°C por 30 minutos sob agitação de 200 rpm. Transcorrido este tempo, 7 mL de dimetilsulfóxido (DMSO) foram

adicionados e o material submetido à nova incubação a 42°C por 20 minutos. Após esta incubação, o material foi incubado em gelo por 10 minutos e centrifugado à temperatura ambiente por 5 minutos a 1000 g. O sobrenadante foi totalmente descartado, as células ressuspensas em meio YPDA (YPD contendo adenina) em um volume final de 11 mL e plaqueadas em meio YNB sólido suplementado com adenina e contendo 30, 10 e 0 mM de 3-aminotriazol (3-AT), para as iscas hINT6, PCI e hINT6T22, respectivamente. O 3-AT atua como um inibidor competitivo da enzima codificada pelo gene *HIS3* e seu uso se justifica pelo fato das iscas hINT6 e PCI ativarem este gene repórter na ausência de interação. Um controle negativo (sem a adição da biblioteca de cDNA) e um controle positivo (a mesma mistura de células plaqueada em meio YNB suplementado com adenina e histidina) foram feitos em paralelo. As placas foram incubadas em estufa a 30°C por até 6 dias. As colônias que cresceram entre o 3° e 6° dias foram transferidas para meio de cultura sólido (YNB suplementado com adenina e histidina) para posterior teste de β -galactosidase.

Ensaio de β -galactosidase

Para os ensaios de β -galactosidase em membrana de nitrocelulose (Bartel & Fields, 1995), as células foram crescidas em meio sólido YNB suplementado com adenina e histidina a 30°C por 2 dias. Uma membrana de nitrocelulose foi colocada sobre a placa até que as colônias fossem adsorvidas e, posteriormente, mergulhada em nitrogênio líquido por alguns segundos. A membrana foi cuidadosamente removida e colocada em contato com papel de filtro embebido em uma solução de tampão Z (87 mM Na₂HPO₄, 47 mM NaH₂PO₄, 19 mM KCl, 1,38 mM MgSO₄, pH 7,0, 0,27% β -mercaptoetanol) contendo 1 mg/ml de X-gal (5-bromo-4-cloro- β -D-galactopiranosídeo). A membrana foi incubada em estufa a 30°C por alguns minutos até o aparecimento de coloração azul nas colônias.

Ensaio quantitativo da atividade de β -galactosidase (Bartel & Fields, 1995) também foram realizados. As colônias de leveduras foram inoculadas em meio mínimo líquido por 16 horas com os aminoácidos apropriados. A OD₆₀₀ foi medida e quantidades iguais de células foram usadas para a realização do ensaio. As células foram centrifugadas e ressuspensas 1,0 mL de tampão Z, contendo 0,004% de duodecil sulfato de sódio (SDS) e 2,0 mM de o-nitrofenil- β -D-galactopiranosídeo (ONPG). Foram adicionados 10

μ L de clorofórmio às amostras e, posteriormente, as mesmas foram homogeneizadas sob forte agitação e incubadas a 30°C até o aparecimento de coloração amarela. A absorbância foi medida a 420 nm em espectrofotômetro.

Teste de crescimento das colônias em placas com 50 mM de 3-AT

Após o isolamento das colônias das placas de triagem do duplo-híbrido, foi realizado um teste para observação de seus crescimentos visando uma seleção mais rigorosa dos clones positivos. Para isto, as colônias foram estriadas em meio YNB suplementado com adenina e contendo 50 mM de 3-AT por 2 dias a 30°C e o crescimento analisado.

Cura dos vetores pBTM-INT6 e pBTM-PCI

A cura de um vetor consiste em sua perda induzida pela levedura através da adição em cultura do suprimento que é a marca seletiva do vetor, neste caso, o triptofano. Não há pressão seletiva para a manutenção do plasmídeo pela célula e com o decorrer das gerações menos células o conterão. Após 4 passagens em meio líquido YNB contendo triptofano, o teste final para verificar a perda dos vetores foi realizado em placas de meio YNB sólido sem este aminoácido. As células foram estriadas também em placas réplicas suplementadas com triptofano para seu isolamento. Assim é possível a obtenção de clones que contenham apenas o vetor pACT2 com o cDNA de interesse.

Isolamento do DNA de levedura

O isolamento do DNA de levedura (Sherman *et al.*, 1986) foi feito a partir de um inóculo em 5 mL de meio YNB, suplementado com os aminoácidos adequados. As células foram crescidas por 16 horas a 30°C e, posteriormente, centrifugadas por 5 minutos a 1000 g. O sobrenadante foi descartado e as células ressuspensas em 1 mL de uma solução de sorbitol 1 M. O material foi novamente centrifugado e o sedimento ressuspenso em 800 μ L de tampão de esferoplasto (0,9 M sorbitol; 50 mM tampão fosfato pH 7,5 e 14 mM β -mercaptoetanol). A este volume foram adicionados 50 μ L de litase (5 mg/mL) com posterior incubação a 37°C até a formação completa de esferoplastos. O material foi centrifugado à temperatura ambiente por 5 minutos a 1000 g e o precipitado homogeneizado em 500 μ L de solução de lise (50 mM Tris-HCl e 20 mM EDTA), com adição de 50 μ L de SDS 10% e 2 μ L de dietilpirocarbonato (DEPC). O material foi

homogeneizado e incubado por 20 minutos a 65°C e posteriormente resfriado em gelo. Foram adicionados 200 µL de uma solução de 5 M de acetato de sódio sendo a amostra incubada em gelo por 60 minutos e centrifugada à temperatura ambiente por 10 minutos a 14.000 g. Posteriormente, 750 µL de isopropanol foram adicionados ao sobrenadante para precipitação do DNA e uma centrifugação por 10 minutos a 14.000 g foi realizada. O DNA foi lavado com etanol 70% e 100% e dissolvido em volume apropriado de TE 1X.

Sequenciamento de DNA e análise das seqüências

O vetor pACT2 contendo o cDNA da biblioteca foi isolado através da extração do DNA da levedura, inserido por eletroporação (Ausubel *et al.*, 1998) na cepa de *E. coli* DH5α e posteriormente preparado para o sequenciamento, realizado com o sistema Big Dye, através de um sequenciador automático modelo DNS ABI PRISM 377 Genetic Analyser (Applied Biosystems). As seqüências obtidas foram comparadas às seqüências do GenBank-NCBI (<http://www.ncbi.nlm.nih.gov/>) através do programa BLAST (Altschul *et al.*, 1990). Análises posteriores foram feitas utilizando os programas ExPASy (Gasteiger *et al.*, 2003) e SMART (Schultz *et al.*, 1998, 2000).

Ensaios de interação *in vitro*

Diversos ensaios de interação *in vitro* foram realizados. As proteínas NY-REN-21, IkBL, alfa4 e 14-3-3τ foram clonadas em fusão com cauda de histidina no vetor pET-28a (Novagen), expressas em *E. coli* e imobilizadas, de acordo com protocolos padrões, em coluna de afinidade Ni-NTA (Qiagen). As proteínas imobilizadas nesta resina foram incubadas com a proteína INT6 em fusão com proteína A, GST e traduzida *in vitro* (Promega, *TnT Quick coupled transcription/translation systems*). Após o tempo de incubação, que variou de 1 hora a 4 horas, a resina foi lavada com diferentes combinações de tampões (modificações na concentração de sal e agentes adstrigentes) e as proteínas eluídas no mesmo tampão contendo 200 mM de imidazol. Foi testada também a imobilização da proteína de fusão GST-INT6 na resina Glutathione Sepharose 4B (Amersham Biosciences) e proteínaA-INT6 na resina IgG Sepharose (Amersham Biosciences) e posterior incubação com as proteínas contendo cauda de histidina. Nestes casos, a eluição foi

realizada em tampão contendo 10 mM de glutathione reduzida e 0,5 M de ácido acético pH 3,4, respectivamente. O resultado foi analisado por *immunoblot* com os anticorpos adequados, com exceção do teste realizado com as proteínas traduzidas *in vitro*, cuja análise foi feita por autorradiograma.

RESULTADOS E DISCUSSÃO

Visando obter dados abrangentes sobre os ligantes protéicos da proteína humana INT6, foram realizadas três triagens pelo método do duplo-híbrido de levedura. Inicialmente o cDNA completo da proteína INT6 e o domínio PCI foram clonados em fusão ao domínio de ligação ao DNA *lexA* de *E. coli* no vetor pBTM116. Estas triagens foram de difícil interpretação, uma vez que as iscas usadas ativavam os genes repórteres do sistema, uma ampla gama de proteínas com diferentes funções foi isolada e muitas dificuldades técnicas foram encontradas na tentativa de confirmação *in vitro* destas interações. Uma nova triagem usando como isca a forma truncada hINT6T22 da proteína INT6 foi realizada, visto que esta proteína não ativou os genes repórteres *HIS3* e *lacZ*.

A triagem pelo método do duplo-híbrido

As triagens para o isolamento de proteínas que interagem com a hINT6, com o domínio PCI e com a forma truncada hINT6T22, foram realizados utilizando-se uma biblioteca de cDNA de cérebro fetal humano (Clontech). Foram obtidos $5,0 \times 10^5$ clones transformantes com a proteína hINT6 como isca e selecionados 316 colônias das placas com 30 mM de 3-AT. Para o domínio PCI foi realizada uma triagem menor, onde $4,0 \times 10^4$ transformantes foram obtidos, dos quais 99 apresentaram crescimento nas placas contendo 10 mM de 3-AT. A seleção das colônias foi realizada até o 6º dia de incubação das placas a 30°C. Desta forma, foram obtidos 39 clones no 4º dia, 123 no 5º e 154 no 6º, na triagem com a proteína hINT6 como isca (identificados pela letra "I"), e 4 clones no 3º dia, 50 no 4º, 28 no 5º e 17 no 6º com o domínio PCI (identificados pela letra "P"). Na triagem para

o isolamento de proteínas que interagem com a forma hINT6T22, foram obtidos $2,9 \times 10^7$ clones transformantes e selecionados 269 das placas de transformação. Dentre estes, 91 foram obtidos no 3º dia de incubação das placas, 98 no 4º dia e 80 nos 5º e 6º dias.

A proteína hINT6 e domínio PCI como iscas:

A presença de um segundo gene marcador na cepa L40 é importante para diminuir o número de interações falso positivas. Após a seleção dos clones transformados para o marcador *HIS3*, os mesmos foram testados em relação ao 2º gene marcador: o gene *lacZ*. A transcrição deste gene codifica a enzima β -galactosidase que confere coloração azul às colônias, na presença do substrato X-gal. Entretanto, com relação à proteína hINT6, este teste feito em filtros de nitrocelulose não se mostrou conclusivo, pois todas as colônias testadas mostraram coloração azul (Figura 1A). Para interações com o domínio PCI foi possível tanto a seleção como a exclusão de alguns clones testados com o substrato X-gal. Aqueles que mostraram coloração azul mais rapidamente e de forma mais intensa após incubação (total de 16) foram selecionados para o isolamento do vetor pACT2. Dezoito clones não se mostraram azuis após o teste, demonstrando que não havia interação entre a proteína codificada pelo cDNA do vetor pACT2 e o domínio PCI, desta forma, foram excluídos de posterior análise (Figura 1B). Os demais clones num total de 65 se mostraram azuis, porém com tempo maior de incubação com o substrato. A cepa L40 transformada com os vetores pBTM-INT6 e pGAD424 (Clontech) foi usada como controle negativo (C-) e como controle positivo foram usados os vetores pBTM-Nip7 e pACT-Nop8 (Zanchin & Goldfarb, 1999).

Para a redução do número de clones falso-positivos destas triagens, foi necessária a realização de um teste quantitativo da atividade enzimática de β -galactosidase, com o substrato ONPG. Este substrato é solúvel em água e apresenta coloração amarela quando hidrolizado pela enzima β -galactosidase, de forma que a intensidade de coloração pode ser medida em espectrofotômetro a 420 nm.

Foram testados 169 clones obtidos da triagem com a proteína hINT6 como isca e os 65 clones isolados com o domínio PCI que não foram diretamente selecionados nem excluídos pelo teste em membrana de nitrocelulose. Além deste ensaio, os clones obtidos na triagem com a proteína hINT6 foram submetidos a um teste para a observação de seus crescimentos em placas contendo 50 mM de 3-AT, visando uma seleção mais rigorosa. O crescimento das colônias foi anotado para posterior comparação.

Foram selecionados 88 clones (68 que possivelmente interagem com a proteína hINT6 e 20 com o domínio PCI) para identificação do cDNA contido no vetor pACT2 por seqüenciamento. Foram escolhidos aqueles clones que apresentaram bom crescimento nas placas com 50 mM de 3-AT e coloração mais intensa nos teste com β -galactosidase.

Os vetores pBTM-INT6 e pBTM-PCI foram eliminados da cepa L40, através do método da cura das leveduras. Este procedimento foi necessário uma vez que os plasmídeos pBTM116 e pACT2 apresentam a mesma marca de resistência a antibiótico, ampicilina, o que impossibilitaria a seleção do vetor pACT2 em *E. coli* para posterior seqüenciamento do cDNA da biblioteca. Desta forma, após a cura, o vetor pACT2 contendo o cDNA da biblioteca foi isolado através da extração do DNA da levedura, transformado por eletroporação na cepa de *E. coli* DH5 α e preparado para o seqüenciamento. Todas as seqüências foram submetidas à comparação no GenBank (NCBI) utilizando o programa BLAST (Altschul *et al.*, 1990).

Os resultados das análises das seqüências estão descritos na tabela 2. Foram encontradas 12 seqüências de DNA cujo mRNA não foi identificado, 5 seqüências de mRNA cuja proteína não é conhecida, 15 proteínas ainda não caracterizadas e 37 proteínas diferentes já caracterizadas. Dentre as proteínas encontradas, caracterizadas ou não, aquelas que estavam *in frame* com a seqüência do domínio de ativação do vetor pACT2, foram analisadas através do programa SMART (Schultz *et al.*, 1998, 2000) para busca de domínios protéicos conhecidos. Um grande

grupo proteínas (13 ao todo) apresentou domínios que interagem com ácidos nucleicos. Dentre estas, 8 possuem domínios de ligação a DNA (dedo de zinco C_2H_2 , SANT- *SANT* *SWI3*, *ADA2*, *N-CoR* e *TFIIIB*, BRLZ- *Basic Region Leucine Zipper* e PHD- *PHD zinc finger*) e 5 apresentam domínios de interação com RNA (RPR, DEXDc- *DEAD-like helicase superfamily* e ZnFU1- *U1-like zinc-finger*, envolvidos em processamento de RNA, e RRM- *Recognition RNA Motif*). Domínios que atuam em interações entre proteínas como ANK (*Ankyrin repeats*), RING (*RING- Ring finger*) e SCAN (*SRE-ZBP*, *CTfin 51*, *Aw-1* e *Number 18*), também foram encontrados em 7 proteínas.

Apesar das dificuldades técnicas em se identificar as interações falso-positivas, duas proteínas, cuja interação com a INT6 foi comprovada anteriormente, eIF3a/p170 (clone I6.91, Asano *et al.*, 1997) e Ret finger (clone P6.10, Morris-Desbois *et al.*, 1999), foram isoladas. Este resultado foi de grande interesse, pois demonstrou que a triagem realizada foi capaz de detectar interações específicas. O grande número de proteínas diferentes isoladas nestas triagens dificultou muito a interpretação dos resultados. Entretanto, algumas proteínas são de especial interesse por estarem envolvidas em interação com DNA, controle da proliferação celular e transdução de sinais, como as proteínas NY-REN-21, I κ BL, alfa4 e 14-3-3 τ . O antígeno NY-REN-21, identificado em carcinoma de células renais (Scanlan *et al.*, 1999) ainda não tem função esclarecida. Esta proteína é uma possível ortóloga da proteína de camundongo ZFP38, com 82% de identidade de aminoácidos. ZFP38 tem função ativadora da transcrição na porção N-terminal (Chowdhury *et al.*, 1992) e representa um marcador para a linhagem neuronal cerebelar (Yang *et al.*, 1996).

A sequência deduzida de aminoácidos da proteína I κ BL indica que ela pode ser membro da família de proteínas regulatórias I κ B (*inhibitor κ B*, Albertella & Campbell, 1994). Proteínas desta família, das quais as mais caracterizadas são I κ B α e I κ B β , apresentam múltiplas repetições ANK (*ankyrin*) que atuam na interação entre proteínas e são necessárias

para a ligação aos fatores de transcrição NF- κ B (*nuclear factor- κ B*, Hatada *et al.*, 1993).

A proteína alfa4, cuja ortóloga em levedura é a *TAP42*, se associa com a subunidade catalítica da fosfatase PP2A (proteína Ser/Thr fosfatase tipo 2A), PP2Ac (Murata *et al.*, 1997). Esta enzima tem sido relacionada como fator chave de controle em numerosos processos celulares básicos como metabolismo, transcrição e transdução de sinal (Wera & Hemmings, 1995). Em *Saccharomyces cerevisiae*, a associação de *TAP42* e PP2Ac é regulado pelo mecanismo de sinalização alvo da rapamicina (TOR- *target-of-rapamycin*, Thomas & Hall, 1997).

A 14-3-3 τ pertencente à super família de proteínas 14-3-3. Estas proteínas foram altamente conservadas ao longo da evolução e estão envolvidas em diversas funções dentro da célula, entre elas transdução de sinais, controle do ciclo celular (Peng *et al.*, 1997) e apoptose (Masters & Fu, 2001) por ligação a serinas fosforiladas (Muslin *et al.*, 1996).

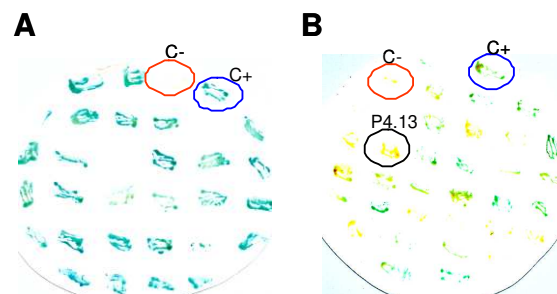


Figura 1: Ensaio de β -galactosidase em membrana de nitrocelulose: Teste realizado com a proteína hINT6 (A) e o domínio PCI (B) como iscas. C+: controle positivo- cepa L40 transformada com os vetores pBTM-Nip7 e pACT-Nop8 (interação positiva); C-: Controle negativo- cepa L40 transformada com os vetores pBTM116 e pGAD424; P4.13: exemplo de clone excluído de posteriores análises (não foi detectada interação entre a proteína expressa pelo vetor pACT2 e o domínio PCI).

Proteína hINT6T22 como isca:

A proteína de fusão *lexA-hINT6T22* não ativa os genes marcadores e a triagem foi realizada em placas sem histidina. O ensaio de β -galactosidase em membrana de nitrocelulose foi realizado com sucesso e se mostrou conclusivo. A cepa L40 co-transformada com os vetores pTL-hINT6T22 e pGAD424 foi usada como controle negativo. Para controle positivo, a cepa L40 foi transformada com os vetores pBTM-Nip7 e pACT-Nop8 (Zanchin & Goldfarb, 1999). Dentre os 269 clones testados, 90 mostraram coloração azul após o tempo de incubação na estufa a 30°C, que variou de 5 minutos a 3,5 horas.

Foram sequenciados 60 cDNAs, que codificam na maioria as proteínas PLZF (*Promyelocytic leukemia zinc finger*), HIRIP3 (*HIRA interacting protein 3*) e histona H3, representando 38%, 13% e 6% de todos os cDNAs sequenciados, respectivamente. Após a realização dos ensaios de ativação dos genes repórteres por estas presas, verificamos que todas foram capazes de ativar os genes *HIS3* e *lacZ* na ausência da isca hINT6T22, o que foi muito surpreendente e revelou que a ativação dos genes repórteres por estas presas não é resultado de uma interação específica. Neste ensaio foram usadas as proteínas ACK-1 (porção C-terminal) e Rab11b como iscas co-transformadas com as presas em questão. Os vetores pTL-CTACK-1 e pTL-Rab11b foram cedidos por Sandra Mara Naressi Scapin. É importante ressaltar que a proteína de fusão *lexA-hINT6T22* foi corretamente expressa na cepa L40 e detectada no extrato da levedura por *immunoblot* contra a proteína *lexA* (Figura 2).

A proteína PLZF pertence à família de proteínas caracterizada pelo domínio N-terminal BTB/POZ (*Broad-Complex*, *Tramtrack*, *Bric-a-brac/Poxvirus and Zinc finger*). Outra característica estrutural da PLZF é a presença em sua porção C-terminal de 9 domínios dedos de zinco C_2H_2 . Os dois primeiros dedos de zinco são dispensáveis para a ligação ao DNA (Li *et al.*, 1997, Sitterlin *et al.*, 1997), embora os outros dedos de zinco da proteína pareçam contribuir para a especificidade da ligação (Ball *et al.*, 1999).

Além disso, um domínio rico em prolina e um domínio ácido são encontrados nesta proteína em sua parte central. A localização em corpos nucleares (Kuken *et al.*, 1997), que são estruturas associadas a um papel central na regulação da transcrição (Zhong *et al.*, 2000), assim como sua baixa regulação na diferenciação das células mielóides, sugerem um papel crucial desta proteína no controle do crescimento celular.

O segundo cDNA mais freqüente nesta triagem foi o que codifica a proteína HIRIP3. Esta proteína foi identificada em uma triagem através do método do duplo-híbrido usando a proteína HIRA (ortóloga humana de proteínas reguladoras da expressão de genes de histonas em *Saccharomyces cerevisiae*) como isca. Além de interagir com a proteína HIRA, HIRIP3 também interage com as histonas H2B e H3. A sequência de aminoácidos desta proteína não revelou ortólogos no banco de dados, nem a presença de domínios protéicos conhecidos (Lorain *et al.*, 1998). Finalmente, a terceira proteína encontrada com freqüência foi a histona H3.

Os resultados obtidos no teste de β -galactosidase somados a dados da literatura sobre a proteína PLZF e sobre a histona H3 fusionada ao domínio de ativação *GAL4* nos levou a conclusão de que estas proteínas não interagem com a proteína hINT6T22 e que os genes repórteres *HIS3* e *lacZ* foram ativados inespecificamente. Foi demonstrado por Sitterlin *et al.* (1997) que a proteína PLZF é capaz de se ligar à sequência de DNA localizada *upstream* dos genes repórteres da cepa L40 ativando a transcrição dos mesmos. Além disso, foi demonstrado que a proteína de fusão *GAL4-histona H3* não é incorporada ao nucleossomo e que é capaz de alterar a expressão gênica de forma específica (Ha *et al.*, 2000).

Embora não tenhamos encontrado na literatura explicação para ativação dos genes repórteres pela proteína HIRIP3, nossos dados mostram claramente que esta proteína ativa a transcrição dos genes *HIS3* e *lacZ*, mesmo quando duas iscas não relacionadas são usadas: C-terminal da proteína ACK-1 (CT-ACK-1) e Rab11b. Este fato ressalta que a ativação dos genes repórteres por esta proteína

não é resultado da interação protéica (Figura 3).

Por esta triagem de duplo-híbrido sustentamos a hipótese de que esta forma truncada da INT6 se comporte como uma forma inativa da proteína perdendo sua função e a capacidade de interação com outras proteínas na célula ou que a proteína não foi corretamente enovelada ou não foi translocada para o núcleo. É importante considerar que o domínio PCI, que se acredita funcionar como um domínio de interação entre proteínas (Yen & Chang, 2000; Kim *et al.*, 2001; Yahalom *et al.*, 2001), não está presente nesta forma truncada o que pode também ter levado a estes resultados.

Testes de interação *in vitro*

Para validar os resultados obtidos através do método do duplo-híbrido de leveduras, é necessária a confirmação *in vitro* das interações. Com este objetivo, selecionamos as proteínas NY-REN-21, I κ BL, alfa 4 e 14-3-3 τ para estes ensaios. A seleção foi realizada com base nos testes de β -galactosidase, nas placas suplementadas com 50 mM de 3-AT, além do tamanho do clone no vetor pACT2 e por se tratarem de proteínas que podem estar envolvidas em transdução de sinais e na proliferação celular. Todas estas proteínas foram clonadas no vetor de expressão pET-28a (Novagem). Após a realização de testes de expressão em *E. coli* verificou-se que as proteínas NY-REN-21, alfa 4 e 14-3-3 τ se encontravam em boa quantidade na fração solúvel do extrato e foram purificadas por cromatografia de afinidade a níquel (dados não mostrados). A proteína I κ BL expressou em baixas quantidades, na fração solúvel do extrato. Após purificação por cromatografia de afinidade a níquel foram observados produtos de menor tamanho provavelmente resultantes de degradação protéica.

A proteína hINT6 não apresenta expressão no extrato solúvel de *E. coli* quando fusionada à cauda de histidina. Em fusão com a proteína GST a expressão na fração solúvel foi muito baixa. Além disso, a purificação foi muito dificultada pela grande quantidade de GST nas

preparações desta proteína de fusão. Aparentemente a fusão GST-INT6 é muito instável sendo a porção da hINT6 degradada rapidamente.

Diversas tentativas para a confirmação *in vitro* das interações entre a proteína hINT6 e as proteínas isoladas pelo duplo-híbrido e clonadas no vetor pET28a foram realizadas. Estes testes foram realizados inicialmente pela imobilização das proteínas com cauda de histidina em coluna de afinidade a níquel e posterior incubação com a proteína hINT6 em fusão com GST ou proteína A ou ainda traduzida *in vitro*.

A situação em que as proteínas de fusão GST-INT6 e proteínaA-INT6 foram imobilizadas e posteriormente incubadas com as outras proteínas também foi testada. Não foi possível comprovar estas interações, uma vez que os controles negativos também apresentavam indícios de interação com as proteínas testadas. O número muito grande de proteínas diferentes encontradas no duplo-híbrido complicou a interpretação dos resultados obtidos que somados à dificuldade de obtenção da proteína hINT6 na forma solúvel resultaram na impossibilidade de confirmação *in vitro* destas interações.

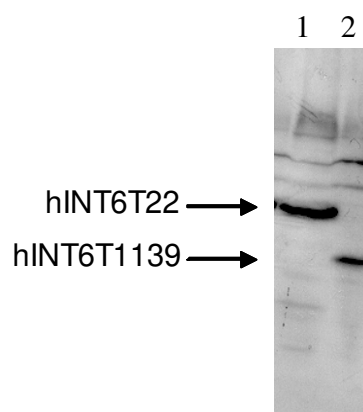


Figura 2: *Immunoblot* realizado para a detecção da proteína de fusão *lexA*-hINT6T22 em extrato da cepa L40 transformada com o vetor pTL-hINT6T22 (1). A cepa L40 transformada com um vetor expressando a forma truncada da hINT6T1139 em fusão com *lexA* foi analisado em paralelo como referência (2).

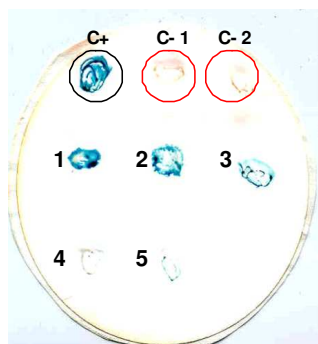


Figura 3: Ensaio de β -galactosidase em membrana de nitrocelulose para teste de ativação do gene *lacZ* pelos diferentes cDNAs da proteína HIRIP3 isolados na triagem usando a hINT6T22 como isca. Neste teste foram usadas as proteínas Rab11b e CT-ACK1 como iscas. C+: controle positivo- pBTM-Nip7/pACT-Nop8; C-1: controle negativo 1- pTL-CT-ACK1/pGAD424; C-2: controle negativo 2- pTL-Rab11b/pGAD424; 1-5: testes; 1- pTL-CT-ACK1/pACT-HIRIP3.1; 2- pTL-CT-ACK1/pACT-HIRIP3.2; 3- pTL-CT-ACK1/pACT-HIRIP3.3; 4- pTL-Rab11b/pACT-HIRIP3.4; 5- pTL-Rab11b/pACT-HIRIP3.5.

CONCLUSÕES

A proteína hINT6, assim como o domínio PCI, quando em fusão ao domínio de ligação ao DNA *lexA* de *E. coli* ativam fortemente os genes repórteres do sistema de duplo-híbrido. Diversas proteínas foram isoladas dificultando a interpretação dos resultados. Apesar disso, duas interações descritas previamente (proteína eIF3a/p170 e Ret finger) foram encontradas com a metodologia usada neste trabalho, o que sugere que apesar do grande número de diferentes proteínas isoladas é possível que existam interações específicas com a hINT6 entre elas. A forma truncada hINT6T22 também foi usada como isca, entretanto, esta triagem apenas detectou interações falso-positivas. Isto sugere que esta forma truncada não foi corretamente enovelada ou não foi translocada para o núcleo ou que o domínio PCI, que está ausente nesta construção, seja imprescindível para as interações da proteína INT6.

Tabela 2: Proteínas obtidas após sequenciamento dos cDNAs da biblioteca de cérebro fetal humano (CLONTECH). Os domínios protéicos encontrados nas proteínas identificadas, assim como a *frame* e o início dos clones no vetor pACT2, também estão apresentados.

Clone	Homologia	Domínios	In frame:	Início do clone no vetor pACT2 (AA)
I4.4, 5.56, I5.101	mRNA DKFZ p761P0615 (a sequência da proteína da proteína ainda não é conhecida)	-----	-----	-----
I4.5	Antígeno se 70-2 de linfoma cutâneo de células T (680 AA)	Motivo de reconhecimento de RNA, (RRM, <i>recognition RNA motif</i>) e sinal de localização nuclear (NLS, <i>Nuclear Localization Signal</i>)	sim	151
I4.7	cDNA do clone RP51119A7	-----	-----	-----
I4.8	RENT1- <i>Regulator of Nonsense Transcript Stability</i> (1118 AA)	DEXDc (DEAD-like helicases superfamily) e AAA (<i>ATPases Associated with diverse cellular Activities</i>)	sim	109
I4.13, I5.1	MPP8- <i>M-phase Phosphoprotein 8</i> (860 AA)	4 ANK (<i>ankyrin</i>) e 1 Organizador e modificador de cromatina (CROMO, <i>chromatin organization modifier</i>)	sim	264
I4.15	mRNA do clone MGC:20115 (a sequência da proteína não é conhecida)	-----	-----	-----
I4.21	Receptor para tirosina fosfatase (979AA)	-----	não	-----
I4.25	Sequência de DNA do cromossomo 17- clone hRPK.180_P	-----	-----	-----

(continua)

(continuação)

Clone	Homologia	Domínios	In frame:	Início do clone no vetor pACT2 (AA)
I4.28, I6.38, I6.42, I6.43, I6.51	ATP sintetase, transportadora de H ⁺ , F1F0 mitocondrial, subunidade d (161 AA)	-----	não	1
I4.29, I5.104	Proteína hipotética MGC 15716 (312 AA)	Dedo de zinco C ₂ H ₂	sim	47
I4.36	Sequência de DNA do clone RP11-141D8	-----	-----	-----
I4.37	Subunidade 2 do fator de splicing 3a (464 AA)	ZnF U1(U1-like zinc finger)	Não definido	-----
I4.40	Sequência de DNA do cromossomo 6 do clone RP11-63719	-----	-----	-----
I4.45	Sequência de DNA do cromossomo 8 do clone RP11-546K22	-----	-----	-----
I4.48, I5.15, I5.30	RAB11b, Ras-related protein (230 AA)	Domínio de ligação a GTP/GTPase	Não/não/sim	1
I4.52	IκBL- I kappa B-like protein (381 AA)	2 ANK (ankyrin)	sim	123
I4.68, I5.95	Proteína hipotética FLJ20531 (474 AA)	Dedo de zinco C ₂ H ₂	sim	90
I5.2	Sequência de cDNA do cromossomo 19 do clone LLNLF-173C4	-----	-----	-----
I5.4	RAB21 GTP-binding protein (225 AA)	-----	não	64
I5.5	Reticulon-2 (206 AA)	2 domínios transmembrana	não	411
I5.6	Sequenciamento não foi informativo	-----	-----	-----
I5.8, I5.85, I6.99	Proteína hipotética KIAA1582 (1309 AA)	Domínio associado a ubiquitina (UBA, <i>Ubiquitin associated domain</i>) e Motivo de reconhecimento de RNA (RRM, <i>recognition RNA motif</i>)	não	1231/1237/1250
I5.10	Mio-inositol-fosfato sintase (558 AA)	-----	sim	302
I5.11	Fosfatidil-inositol fosfato quinase (669 AA)	-----	-----	Homologia fora da sequência codificadora
I5.17	mRNA do clone FLB3401 (a sequência da proteína da proteína ainda não é conhecida)	-----	-----	-----
I5.22	Sequenciamento não foi informativo	-----	-----	-----
I5.23	Sequência de DNA do cromossomo 3 do clone RP11-16M3	-----	-----	-----
I5.24	Sequenciamento não foi informativo	-----	-----	-----
I5.34	Rbx1 – Ring box protein (108 AA)	Domínio dedo de anel (RING finger)	sim	1
I5.40	Fator de <i>splicing</i> 8, rico em arginina e serina (951 AA)	-----	não	912
I5.42	DNA genômico	-----	-----	-----
I5.46	Proteína hipotética KIAA 1623 (912 AA)	4 domínios transmembrana	sim	735
I5.67, I5.68	antígeno NY-REN-21 (473 AA)	SCAN (SRE-ZBP, CTfin 51, Aw-1 e Number), dedos de zinco C ₂ H ₂	sim	52
I5.83	RPIP8 -Rap2-interacting protein (405 AA)	RUN (domínio envolvido em sinalização Ras-like GTPase)		

(continua)

(continuação)

Clone	Homologia	Domínios	In frame:	Início do clone no vetor pACT2 (AA)
I5.100	Antígeno NY-REN-41 reconhecido por anticorpos autólogos em carcinomas de células renais (241AA)	-----	sim	1
I.5.114	RHO-GEF-7 Fator de troca do nucleotídeo guanina (758 AA)	RhoGEF (<i>guanine nucleotide factor for Rho/Rac/Cdc42-like GTPases</i>) e PH (<i>Pleckstrin homology domain</i>)	sim	44
5.115	Proteína hipotética MGC3101 (156 AA)	-----	sim	147
I5.123	mRNA do cDNA DKFZp564M113 (a sequência da proteína da proteína ainda não é conhecida)	-----	-----	-----
I6.8	Homólogo a proteína INT3 (2318 AA)	EGF-CA (<i>Calcium-binding EGF-like domain</i>), NL (domínio encontrado em Notch e Lin-12), repetição ANK (<i>ankyrin</i>), GRAN (<i>granulin</i>)	sim	2078
I6.11	Proteína hipotética MGC13125 (619 AA)	-----	sim	256
I6.12	Proteína hipotética FLJ21022 (305 AA)	-----	sim	154
I6.29	Proteína hipotética XP_049835 (145 AA)	-----	sim	36
I6.44	Regulador de cromatina dependente de actina, associado à matriz relacionado ao complexo SWI/SNF (subfamília c, membro-2) (1214 AA)	-----	não	913
I6.45	Similar à proteína A (313 AA)	transmembrana	sim	230
I6.48	Similar à proteína de ligação do domínio C-terminal da RNA pol II (1268 AA).	RPR (processamento de RNA) e RRM (<i>RNA recognition motif</i>)	sim	1015
I6.49	Proteína guanidilacetato N-metiltransferase (GAMT) (236AA)	-----	não	138
I.6.52	Homóloga a proteína AK 003636 de <i>Mus musculus</i> (335 AA)	-----	sim	18
I.6.59	Proteína tirosina quinase ACK1 (1091AA)	Domínio catalítico de tirosina quinase (Tyrkc, <i>Tyrosine Kinase, catalic domain</i>) e domínio 3 homólogo a Src (SH3, <i>Src homology domain 3</i>)	sim	809
I6.71	Proteína KIAA0147 (1564 AA)	LRR (<i>Leucine-rich repeats</i>), DHR (<i>Dlg homologous region</i>)	sim	1033
I6.77	mRNA do clone MGC3967	-----	-----	-----
I6.91	Subunidade eIF3a/p170 do eIF3 (1382 AA)	PCI (Proteassomo, complexo COP9/signalossomo, fator de iniciação 3)	sim	113
I6.92	Sequência de cDNA do clone BAC15E1	-----	-----	-----
I6.115	DNA do clone BAC RP11-13G14	-----	-----	-----
I6.118	DNA do clone RP5-1070G24	-----	-----	-----

(continua)

(conclusão)

Clone	Homologia	Domínios	In frame:	Início do clone no vetor pACT2 (AA)
P3.2	Clusterina/TRPM-2 (<i>testosterone-repressed prostate message</i> , 449 AA)	CLa (<i>Clusterin alpha chain</i>), CLb (<i>Clusterin beta chain</i>)	sim	285
P3.3	Regulador de cromatina dependente de actina, associado à matriz relacionado ao complexo SWI/SNF (subfamília f, membro-1, 1902 AA)	-----	não	1792
P3.4	Homóloga a CDC5 (<i>cell division cycle 5</i>) (802 AA)	SANT (<i>SWI3, ADA2, N-CoR and TFIIIB</i>) DNA-binding domains)	sim	715
P4.15, P4.35	14-3-3, isoforma τ (245 AA)	14_3_3	sim	65/78
P4.23	Proteína ligadora de imuglobulina 1, alfa 4, similar a proteína TAP42 de levedura (340AA)	-----	sim	23
P4.27	Proteína hipotética (163AA)	-----	sim	6
P4.39	Seqüenciamento não foi informativo	-----	-----	-----
P4.40	CREB (<i>Cyclic AMP response element binding protein</i> , 327 AA)	BRLZ (<i>Basic Region Leucin Zipper</i>)	-----	Homologia fora da seqüência codificadora
P4.41	Proteína hipotética FLJ23445 (204 AA)	BRLZ (<i>Basic Region Leucin Zipper</i>)	sim	1
P4.43	Proteína KIAA0376 (952 AA)	Domínio homólogo à calponina, (CH <i>Calponin homolog domain</i>), BRLZ (<i>basic region leucin zipper</i>), MA (<i>Methyl accepting chemotaxis-like domains</i>)	sim	487
P4.45	Sintaxina 18 (335 AA)	1 domínio transmembrana	sim	1
P4.46	Seqüência de DNA do cromossomo 19 do clone RP11-637019	-----	-----	-----
P4.47	Proteína hipotética FLJ20014 (223 AA)	-----	sim	181
P4.48	Similar à cadeia pesada da ferritina (152 AA)	-----	não	58
P4.49	Proteína associada à proteína similar a FKBP (<i>FK-binding protein</i> , 596AA)	-----	sim	547
P5.2	Cadeia alfa da tubulina (451 AA)	-----	sim	1
P5.28	Proteína KIAA0804 (1211 AA)	Domínio dedo de anel (<i>RING finger</i>), PHD (<i>PHD zinc finger</i>)	sim	693
P5.39	Isoforma 1 da proteína A (284 AA)	-----	não	119
P6.10	Isoforma alfa da proteína Ret finger (513 AA)	Domínio dedo de anel (<i>RING finger</i>), BBOX (<i>B-Box-type zinc-finger</i>), PRY (associado com o domínio SPRY, de função desconhecida), SPRY (<i>Domain in Spla and the Ryanodine Receptor</i>)	sim	208

AA: aminoácido; -----: domínios protéticos ausentes ou não investigados.

REFERÊNCIAS

- Akiyoshi, Y., Clayton, J., Phan, L., Yamamoto, M., Hinnebusch, A. G., Watanabe, Y., Asano, K. 2001. Fission yeast homolog of murine Int-6 protein, encoded by mouse mammary tumor virus integration site, is associated with the conserved core subunits of eukaryotic translation initiation factor 3. *J Biol Chem*, 276:10056-10062.
- Albertella, M. R. & Campbell, R. D. 1994. Characterization of a novel gene in the human major histocompatibility complex that encodes a potential new member of the I kappa B family of proteins. *Hum Mol Genet*, 3:793-799.
- Altschul, S. F., Gish, W., Millar, W., Myers, E. W., Lipman, D. J. 1990. Basic local alignment search tool. *J Mol Biol*, 215:403-410.
- Asano, K., Merrick, W. C., Hershey, J. W. B. 1997. The translation initiation factor eIF3-p48 subunit is encoded by int-6, a site of frequent integration by the mouse mammary tumor virus genome. *J Biol Chem*, 272:23477-23480.
- Ausubel, F. M., Brent, R., Kingston, R., Moore, D. D., Seidman, J. G., Smith, J. A., Struhl, K. 1998. Current Protocols in Molecular Biology. John Wiley & Sons, New York, USA.
- Ball, H. J., Melnick, A., Shakhovich, R., Kohanski, R. A., Licht, J. D. 1999. The promyelocytic leukemia zinc finger (PLZF) protein binds DNA in a high molecular weight complex associated with cdc2 kinase. *Nucleic Acids Res*, 27:4106-4113.
- Bandyopadhyay, A., Matsumoto, T., Maitra, U. 2000. Fission yeast Int6 is not essential for global translation initiation, but deletion of *int6+* causes hypersensitivity to caffeine and affects spore formation. *Mol Biol Cell*, 11:4005-4018.
- Bandyopadhyay, A., Lakshmanan, V., Matsumoto, T., Chang, E. C., Maitra, U. 2002. Moe1 and spInt6, the fission yeast homologues of mammalian translation initiation factor 3 subunits p66 (eIF3d) and p48 (eIF3e), respectively, are required for stable association of eIF3 subunits. *J Biol Chem*, 277:2360-2367.
- Bartel, P. L., Chien, C., Sternglanz, R., Fields, S. 1993. Using the two-hybrid system to detect protein-protein interactions. D. A. Hartleu (ed). Cellular interactions in development: a practical approach. 153-179. IRL Press, Oxford.
- Bartel, P. L. & Fields, S. 1995. Analyzing protein-protein interactions using two-hybrid system. *Methods in Enzymology*, 254:241-263.
- Browning, K. S., Gallie, D. R., Hershey, J. W., Hinnebusch, A. G., Maitra, U., Merrick, W. C., Norbury, C. 2001. Unified nomenclature for the subunits of eukaryotic initiation factor 3. *Trends Biochem Sci*, 26:284.
- Carneiro, F. R., Silva, T. C., Alves, A. C., Haline-Vaz, T., Gozzo, F. C., Zanchin, N. I. 2006. Spectroscopic characterization of the tumor antigen NY-REN-21 and identification of heterodimer formation with SCAND1. *Biochem Biophys Res Commun*, 343:260-268.
- Chowdhury, K., Goulding, M., Walther, C., Imai, K., Fickenscher, H. 1992. The ubiquitous transactivator Zfp-38 is upregulated during spermatogenesis with differential transcription. *Mech Dev*, 39:129-142.
- Desbois, C., Rousset, R., Bantignies, F., Jalinot, P. 1996. Exclusion of Int-6 from the PML nuclear bodies by binding to the HTLV-I Tax oncoprotein. *Science*, 273:951-953.
- Dickson, C., Smith, R., Brookes, S., Peters, G. 1984. Tumorigenesis by mouse mammary tumor virus: proviral activation of a cellular gene in the common integration region int-2. *Cell*, 37:529-536.
- Fields, S. & Song, O. 1989. A novel genetic system to detect protein-protein interaction. *Nature*, 340:245-246.
- Gallahan, D. & Callahan, R. 1987. Mammary tumorigenesis in feral mice: identification of a new int locus in mouse mammary tumor virus (Czech II)-induced mammary tumors. *J Virol*, 61:66-74.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., Bairoch, A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, 31:3784-3788.
- Gietz, R. G. & Sugino, A. 1988. New yeast-Escherichia coli shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. *Gene*, 74:527-534.
- Glickman, M. H., Rubin, D. M., Coux, O., Wefes, I., Pfeifer, G., Cjeda, Z., Baumeister, W., Fried, V. A., Finley, D. 1998. A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-Signosome and eIF3. *Cell*, 94:615-623.
- Guo, J. & Sen, G. C. 2000. Characterization of the interaction between the interferon-induced protein P56 and the Int-6 protein encoded by a locus of insertion of the mouse mammary tumor virus. *J Virol*, 74:1892-1899.
- Guo, J., Hui, D. J., Merrick, W. C., Sen, G. C. 2000. A new pathway of translational regulation mediated by eukaryotic initiation factor 3. *EMBO*, 19:6891-6899.
- Ha, N., Hellauer, K., Turcotte, B. 2000. Fusions with histone H3 result in highly specific alteration of gene expression. *Nucleic Acids Res*, 28:1026-1035.

- Hatada, E. N., Naumann, M., Scheidereit, C. 1993. Common structural constituents confer I kappa B activity to NF-kappa B p105 and I kappa B/MAD-3. *EMBO J*, 12:2781-2788.
- Hofmann, K. & Bucher, P. 1998. The PCI domain: a common theme in three multiprotein complexes. *TIBS*, 23:204-205.
- Hollenberg, S. M., Sternglanz, R., Cheng, P. F. and Weintraub, H. 1995. Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system. *Mol Cell Biol*, 15:3813-3822.
- Kim, T., Hofmann, K., von Armin, A. G., Chamovitz, D. A. 2001. PCI complexes: pretty complex interactions in diverse signaling pathways. *Trends Plant Sci*, 6:379-386.
- Kuken, M. H., Reid, A., Quignon, F., Chelbi-Alix, M. K., Davies, J. M., Kabarowski, J. H., Zhu, J., Dong, S., Chen, S., Chen, Z., Tan, C. C., Licht, J., Waxman, S., de The, H., Zelent, A. 1997. Leukemia-associated retinoic acid receptor alpha fusion partners, PML and PLZF, heterodimerize and colocalize to nuclear bodies. *Proc Natl Acad Sci*, 94:10255-10260.
- Li, J. Y., English, M. A., Ball, H. J., Yeyati, P. L., Waxman, S., Licht, J. D. 1997. Sequence-specific DNA binding and transcriptional regulation by the promyelocytic leukemia zinc finger protein. *J Biol Chem*, 272:22447-22455.
- Lorain, S., Quivy, J. P., Monier-Gavelle, F., Scamps, C., Lecluse, Y., Almouzni, G., Lipinski, M. 1998. Core histones and HIRIP3, a novel histone-binding protein, directly interact with WD repeat protein HIRA. *Mol Cell Biol*, 18:5546-56.
- Marchetti, A., Batitta, F., Miyazaki, S., Gallahan, D., Smith, G. H., Gallahan, R. 1995. Int-6, a highly conserved, widely expressed gene, is mutated by mouse mammary tumor virus in mammary preneoplasia. *J Virol*, 69:1932-1938.
- Masters, S. C & Fu, H. 2001. 14-3-3 proteins mediate an essential anti-apoptotic signal. *J Biol Chem*, 276:45193-45200.
- Mayeur, G. L. & Hershey, J. W. 2002. Malignant transformation by the eukaryotic translation initiation factor 3 subunit p48(eIF3e). *FEBS*, 514:49-54.
- Miyazaki, S., Imatani, A., Ballard, L., Marchetti, A., Buttitta, F., Albertsen, H., Nevanlinna, H., A., Gallahan, D., Callahan, R. 1997. The chromosome location of the human homolog of the mouse mammary tumor associated gene INT6 and its status in human breast carcinomas. *Genomics*, 46:155-158.
- Morris-Desbois, C., Bochar, V., Reynaud, C., Jalinot, P. 1999. Interaction between the Ret finger protein and the Int-6 gene product and co-localisation into nuclear bodies. *J Cell Sci*, 112:3331-3342.
- Morris-Desbois, C., Rety, S., Myriam F., Garin, J., Jalinot, P. 2001. The human protein HSPC021 interacts with Int-6 and is associated with Eukaryotic translation initiation factor 3. *J Biol Chem*, 276:45988-45995.
- Murata, K., Wu, J., Brautigan, D. L. 1997. B cell receptor-associated protein alpha4 displays rapamycin-sensitive binding directly to the catalytic subunit of protein phosphatase 2A. *Proc Natl Acad Sci*, 94:10624-10629.
- Muslin, A. J., Tanner, J. W., Allen, P. M., Shaw, A. S. 1996. Interaction of 14_3_3 with signaling proteins is mediated by the recognition of phosphoserine. *Cell*, 84: 889-897.
- Nusse, R. & Varmus, H. E. 1982. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell*, 31:99-109.
- Peng, C. Y., Graves, P. R., Thoma, R. S., Wu, Z., Shaw, A. S., Piwinica-Worms, H. 1997. Mitotic and G2 checkpoint control: regulation of 14_3_3 protein binding by phosphorylation of Cdc25C on serine-216. *Science*, 277:1501-1505.
- Peters, G., Brookes, S., Smith, R., Placzek, M., Dickson, C. 1989. The mouse homolog of the hst/k-FGF gene is adjacent to int-2 and is activated by proviral insertion in some virally induced mammary tumors. *Proc Natl Acad Sci U S A*, 86:5678-5682.
- Rasmussen, S. B., Kordon, E., Callahan, R., Smith, G. H. 2001. Evidence for the transforming activity of a truncated int6 gene, *in vitro*. *Oncogene*, 20:5291-5301.
- Scanlan, M. J., Gordan, J. D., Williamson, B., Stockert, E., Bander, N. H., Jongeneel, V., Gure, A. O., Jager, D., Jager, E., Knuth, A., Chen, Y. T., Old, L. J. 1999. Antigens recognized by autologous antibody in patients with renal-cell carcinoma. *Int J Cancer*, 83:456-464.
- Scheel, H. & Hofmann, K. 2005. Prediction of a common structural scaffold for proteasome lid, COP9-signalosome and eIF3 complexes. *BMC Bioinformatics*, 6:71.
- Schultz, J., Milpetz, F., Bork, P., Ponting, C. P. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci*, 95:5857-5864.
- Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., Bork, P. 2000. SMART: a web-based tool for study of genetically mobile domains. *Nucleic Acids Res*, 28:231-234.

- Sherman, F., Fink, G. R., Hicks, J. B. 1986. Laboratory Course Manual for Methods in Yeast Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sitterlin, D., Tiollais, P., Transy, C. 1997. The RAR alpha-PLZF chimera associated with Acute Promyelocytic Leukemia has retained a sequence-specific DNA-binding domain. *Oncogene*, 14:1067-1074.
- Tekmal, R. R. & Keshava, N. 1997. Role of MMTV integration locus cellular genes in breast cancer. *Front Biosci*, 2:519-526.
- Thomas, G. & Hall, M. N. 1997. TOR signalling and control of cell growth. *Curr Opin Cell Biol*, 9:782-787.
- Wera, S. & Hemmings, B. A. 1995. Serine/threonine protein phosphatases. *Biochem*, 311:17-29
- Yahalom, A., Kim, T. H., Winter, E., Karniol, B., von Arnim, A. G., Chamovitz, D. A. 2001. Arabidopsis eIF3e (INT-6) associates with both eIF3c and the COP9 signalosome subunit CSN7. *J Biol Chem*, 276: 334-340.
- Yang, X. W., Zhong, R., Heintz, N. 1996. Granule cell specification in the developing mouse brain as defined by expression of the zinc finger transcription factor RU49. *Development* 122:555-566.
- Yen, H. C. & Chang, E. C. 2000. Yin6, a fission yeast Int6 homolog, complexes with Moe 1 and plays a role in chromosome segregation. *Cell Biology*, 97:14370-14375.
- Yen, H. C., Gordon, C., Chang, E. C. 2003. *Schizosaccharomyces pombe* Int6 and Ras homologs regulate cell division and mitotic fidelity via the proteasome. *Cell*, 112:207-217.
- Zachin, N. I. & Goldfarb, D. S. 1999. Nip-7p interacts with Nop8p, an essential nucleolar protein required for 60S ribosome biogenesis, and the exosome subunit Rrp43p. *Mol Cell Biol*, 19:1518-1525.
- Zhong, S., Salomoni, P., Pandolfi, P. P. 2000. The transcriptional role of PML and the nuclear body. *Nat Cell Biol*, 2: E85-E90.

3.2- Artigo I

Identification and characterization of a proteolysis-resistant fragment containing the PCI domain in *Arabidopsis thaliana* INT6/eIF3e translation factor

Marcelo J. Murai, Flávia R. G. Carneiro, Fábio C. Gozzo, Daniela F. Ierardi, Thelma A. Pertinhez and Nilson I.T. Zanchin

ORIGINAL ARTICLE

Identification and Characterization of a Proteolysis-Resistant Fragment Containing the PCI Domain in the *Arabidopsis thaliana* INT6/eIF3e Translation Factor

Marcelo J. Murai, Flávia R. G. Carneiro, Fabio C. Gozzo, Daniela F. Ierardi, Thelma A. Pertinhez, and Nilson I. T. Zanchin*

Center for Structural Molecular Biology, Brazilian Synchrotron Light Laboratory-LNLS, Campinas-SP, Brazil

Abstract

The PCI domain comprises approx 200 amino acids and is found in subunits of the eukaryotic translation initiation factor 3 (eIF3), the 26S proteasome and the COP9/signalosome complexes. The PCI domain is involved in protein-protein interaction, and mouse INT6 truncated proteins lacking the PCI domain show cell malignant-transforming activity. In this work, the *Arabidopsis thaliana* INT6/eIF3e (AtINT6) protein was dissected using limited proteolysis, and a protease-resistant fragment containing the PCI domain was identified. Based on mass spectrometry analyses of the protease-resistant fragments and on secondary structure prediction, AtINT6-truncated proteins were cloned and expressed in *Escherichia coli*. Stability studies using thermal unfolding followed by circular dichroism revealed a midpoint transition temperature of 44°C for the full-length AtINT6 protein, whereas the truncated proteins comprising residues 125-415 (AtINT6TR2) and 172-415 (AtINT6TR3) showed transition temperatures of 49 and 58°C, respectively. AtINT6TR3 contains the PCI domain with additional amino acids at the N and C termini. It shows high solubility, and together with the high thermal stability, should facilitate further characterization of the PCI domain structure, which is important to understand its function in protein-protein interaction.

Index Entries: PCI domain; eIF3; INT6; limited proteolysis; circular dichroism.

INTRODUCTION

Sequence alignment identified the PCI domain in the subunits of three large protein complexes: the 26S proteasome, the plant photomorphogenic regulatory complex (COP9), and the translation initiation factor 3 complex (eIF3) (1). It has been predicted to be a 200-amino acid α -helical domain involved in protein-protein interaction (1,2).

INT6 is a PCI domain-containing protein first identified in mice, being encoded by a gene located in a frequent genome integration site of the mouse mammary tumor virus (MMTV) (3). Studies with truncated proteins generated by MMTV integration in the mouse

genome indicated that INT6 might be involved in the control of cell proliferation (4,5). Interestingly, in this case, these proteins did not contain the PCI domain. INT6 was subsequently shown to copurify with the eIF3 complex isolated from both reticulocyte lysates and human HeLa cells (6). In addition, it has been suggested to act in translational repression via interaction with the P56 interferon-induced protein (7,8). INT6 interacts also with the human type I T-cell leukemia virus Tax oncoprotein (9) and with the Ret finger protein 1 (10). INT6 interaction with the Tax leads to subcellular redistribution (9) that probably affects its function. Indeed, under normal conditions INT6 shows a cell cycle-dependent localization with a preference for the nuclear compartment (11). In *Arabidopsis thaliana*, the INT6 ortholog (AtINT6) has been found preferentially in the nucleus and interacts directly with the *c* subunit of eIF3 and

* Author to whom all correspondence and reprint requests should be addressed. E-mail: zanchin@lnls.br

with the CSN7 subunit of the COP9 signalosome (12). The COP9 complex is a suppressor of light-dependent development; nonetheless, equivalent complexes are found in eukaryotic organisms with one exception, the yeast *Saccharomyces cerevisiae*. Therefore, protein interaction studies in *A. thaliana* provided evidence that the nuclear role of INT6 is associated with the function of the COP9 signalosome.

INT6 has many interacting partners and its interaction activity has been assigned to the PCI domain that is found in the C-terminal half of the protein (2,12). However, it is unclear how this domain can recognize multiple proteins, and the characterization of its structure should provide relevant information about its mechanism of action. Recently, two nuclear magnetic resonance structures of partial PCI domains have been deposited in the Protein Data Bank (PDB) by S. Suzuki and coworkers for the mouse hypothetical protein Aah51541 and the COP9 complex subunit 4 (PDB accession no. 1WI9 and 1UFM, respectively). These partial domains correspond to approximately one-third of the domain identified by sequence alignment (1,2) and do not provide significant information about the PCI domain function. Characterization of the mammalian INT6 is hampered by the insolubility shown by the recombinant protein. These facts have prompted us to select the soluble *A. thaliana* INT6 protein, which shows 62% amino acid similarity to the human counterpart, for limited proteolysis and conformational studies. The protease-resistant fragments were identified by mass spectrometry (MS), and some truncated forms of AtINT6 related to these fragments and containing the PCI domain were constructed. They showed different expression efficiency, solubility, and thermal stability. Most importantly, the results presented in this work evidenced that we have been able to identify a soluble and highly stable AtINT6-truncated protein containing the PCI domain, thus opening the way to the characterization of the PCI structure.

MATERIALS AND METHODS

Plasmid Construction, Protein Expression, and Purification

The cDNA encoding *A. thaliana* INT6 (AtINT6) (National Center for Biotechnology Information accession no. AF285832) was amplified by reverse transcription followed by polymerase chain reaction (PCR) from total RNA by using oligonucleotides ONZ39 (5'-AAAAGGATCCCTAGCGAGTTGCTTGCGCCTG-3') and ONZ41 (5'-GAGAGGATCCCATATGGAGGAAAGCAAACAGAAC-3'). The PCR product was inserted into the *Nde*I and *Bam*HI sites of the expression vector pET28a (Novagen, Madison, WI), producing plasmid

pET-AtINT6. Three truncated forms of AtINT6 were subcloned into the *Nde*I and *Bam*HI sites of pET28a plasmid to express the PCI-containing region of AtINT6. The cDNA region containing amino acids 124–441 was amplified by PCR using oligonucleotides ONZ127 (5'-CCAGCATATGTACCAGATTGGTCCAGACC-3') and ONZ39 to produce vector pET-AtINT6TR1. Vector pET-AtINT6TR2 contains the cDNA region from amino acids 124 to 415, which was PCR-amplified using oligonucleotides ONZ127 and ONZ128 (5'-CTGGATCCTAGTTTATCAACTGCTCATGCAG-3'), and vector pET-AtINT6TR3 contains the cDNA region encoding amino acids 172–415, amplified by PCR using oligonucleotides ONZ235 (5'-GAGTCATATGTGGGGAAAGCTCGCATCTGA-3') and ONZ128. All vectors express proteins with a histidine-tag at the N terminus, and the inserts were verified by DNA sequencing.

Proteins were expressed using *E. coli* strain BL21(DE3) *slyD*⁻, kindly provided by Dr. Ryland Young (Texas A&M University, College Station, TX). Protein expression assays and cell extract preparation were performed as described previously (13). Full-length AtINT6 and AtINT6TR2 proteins were purified by metal-chelating chromatography followed by ion-exchange chromatography on Q-Sepharose (GE Healthcare, Little Chalfont, Buckinghamshire, UK). AtINT6TR3 was purified by a single metal-chelating affinity chromatography step. For affinity purification, the extract from 2 L of LB cultures was incubated with 400 μ L of Ni²⁺-NTA (QIAGEN, Valencia, CA) for 1 h at 4°C; transferred to a spin column, which was washed with 20 column volumes (*cv*) of buffer A (50 mM sodium phosphate, pH 7.2, 100 mM NaCl, 5% glycerol, 1.5 mM imidazole, and 0.5 mM phenylmethylsulfonyl fluoride [PMSF]); and eluted with a linear 0–200 mM imidazole gradient in 45 *cv* of buffer A. Ion-exchange chromatography was performed on a 5-mL Q-Sepharose column (GE Healthcare). The column was washed with 10 mL of a buffer 1 (10 mM sodium phosphate, pH 7.2, 20 mM NaCl, 5% glycerol, and 3.5 mM 2-mercaptoethanol) and eluted with 10 *cv* by using a 0–1000 mM NaCl gradient. Protein samples were concentrated using Ultrafree-4 centrifugal filter devices (Millipore, Billerica, MA), and protein concentration was determined by extinction coefficient at A₂₈₀.

Circular Dichroism Spectroscopy

Circular dichroism (CD) experiments were performed using a Jasco J-810 spectropolarimeter (Jasco, Tokyo, Japan) coupled to a Peltier Jasco PFD-425 system for temperature control. Protein samples were prepared at 4 μ M in 10 mM sodium phosphate buffer at pH 7.2. The spectra were collected using a 1-mm optical path cell in the spectral range of 190–260 nm, and a 1-mm bandwidth with step size of 0.5 nm and scan speed of 50 nm/min. For each measurement, the mean values of four spectra were taken to improve the signal-to-noise ratio. The ellip-

ticity was expressed as the mean residue molar ellipticity (θ) ($\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$). Thermal unfolding experiments were performed by following the loss of secondary structure at 222 nm in the range from 10 to 90°C.

Limited Proteolysis of AtINT6

Sequencing-grade proteases (trypsin, thermolysin, chymotrypsin, papain, and subtilysin) were purchased from Sigma (St. Louis, MO). One hundred-microliter limited proteolysis reactions were performed in 10 mM Tris, pH 7.4, 10 mM NaCl, and 3.5 mM 2-mercaptoethanol at room temperature for up to 4 h. The concentration of AtINT6 was 1 mg/mL, and the enzyme-to-substrate ratio used was 1:100 (w/w). At time intervals, aliquots were removed and boiled in sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) sample buffer to inactivate the proteases. For MS analyses, trypsin was inactivated with 1 mM PMSF. Time-course digestions were analyzed by SDS-PAGE, and the protease-resistant fragments were analyzed by MS.

Mass Spectrometry

For matrix-assisted laser desorption ionization/time of flight (MALDI-TOF) peptide fingerprinting, proteolysis reactions were fractionated on SDS-PAGE, and bands were subjected to in-gel trypsin digestion for 16 h for complete digestion. Samples were desalted with ZipTipC18 (Millipore), according to manufacturer's instructions. Subsequently, 0.5 μL of the sample was spotted with 0.5 μL of matrix solution on a stainless steel sample plate and allowed to dry at room temperature. α -Cyano-4-hydroxycinnamic acid (Sigma) at 10 mg/mL in water/acetonitrile (1:1, v/v) and 0.1% formic acid were used as matrix for ionization. Mass spectra were acquired on a Voyager-DE PRO MALDI-TOF (Applied Biosystems, Foster City, CA) spectrometer. MS using electrospray ionization was performed on a Waters/Micromass Q-ToF Ultima API spectrometer (Waters, Milford, MA) with a nanoflow interface. The digested sample (10 μL) was desalted using a Waters Opti-Pak C18 trap column for 5 min. The sample was eluted from the C18 trap column at a 250 nL/min flow with a mobile phase containing water/acetonitrile (1:1, v/v) and 0.1% formic acid. The instrument conditions were 3 kV for the spray voltage, 100 V for the cone voltage, cone gas at 30 L/h, and source temperature of 100°C. The final spectrum was processed using the MaxEnt1 program (Waters/Micromass).

RESULTS AND DISCUSSION

Recombinant AtINT6 Protein Contains a Protease-Resistant Domain

The *A. thaliana* INT6/eIF3e ortholog (AtINT6) is a 441-amino acid protein sharing 62% amino acid similar-

ity with the human INT6 protein (9). The human INT6 is insoluble when expressed in bacterial systems (data not shown); nonetheless, soluble AtINT6 can be obtained using *E. coli*. Histidine-tagged AtINT6 was purified by metal-chelating affinity and ion-exchange chromatography (Fig. 1, lane I). To determine whether the PCI domain defined by primary structure alignment (1,2) also corresponds to a structural domain, we conducted limited proteolysis experiments by using proteases showing different specificities (trypsin, thermolysin, chymotrypsin, papain, and subtilysin) to digest recombinant AtINT6 (Fig. 1). SDS-PAGE analyses of the protease-resistant products showed that chymotrypsin produced a major fragment of approx 30 kDa (Fig. 1A). Thermolysin and trypsin produced two well-defined bands of approx 25 and 30 kDa and 18 and 30 kDa, respectively (Fig. 1A,B). Although there is accumulation of an approx 30-kDa band, the sample digested with papain occurs as a smear of digested peptides (Fig. 1C). For subtilysin, only a band of approx 18 kDa was detectable at intermediate incubation times (Fig. 1D). That an approx 30-kDa band was generated by limited proteolysis with four of the five proteases tested (chymotrypsin, thermolysin, trypsin, and papain) indicated that the proteolysis-resistant fragment should correspond to the same region of AtINT6.

PCI Domain Is Part of the AtINT6 Protease-Resistant Product

To identify the amino acid sequence of the trypsin- and thermolysin-resistant fragments, we initially performed peptide fingerprinting by using MALDI-TOFMS. The peptides identified for the approx 30-kDa trypsin-resistant domain encompass the region from amino acid 167 to 418 (Fig. 2); however, several of the internal peptides could not be detected. Nevertheless, the predicted molecular mass of this region is 29,623.96 Da and is consistent with the size of the product observed on SDS-PAGE. The peptides identified for the approx 30- and 25-kDa thermolysin fragments contain amino acids 124–385 (predicted molecular mass of 31,056 Da) and amino acids 175–385 (predicted molecular mass of 25,136 Da), respectively. Although the two fragments showed sequence overlapping in the region from 175 to 385 and molecular masses consistent with the bands observed by SDS-PAGE, the peptide fingerprinting strategy turned out to be inconclusive, because several internal peptides could not be detected, thus raising doubts whether the N- and C-terminal ends of the proteolysis products were accurately determined. Therefore, the AtINT6 trypsin-resistant products also were analyzed by MS under electrospray ionization. Analysis with this technique identified a major product of 29,623 Da (Fig. 3), which includes amino acid residues from 167 to 418 because the predicted molecular mass for this region of AtINT6 is

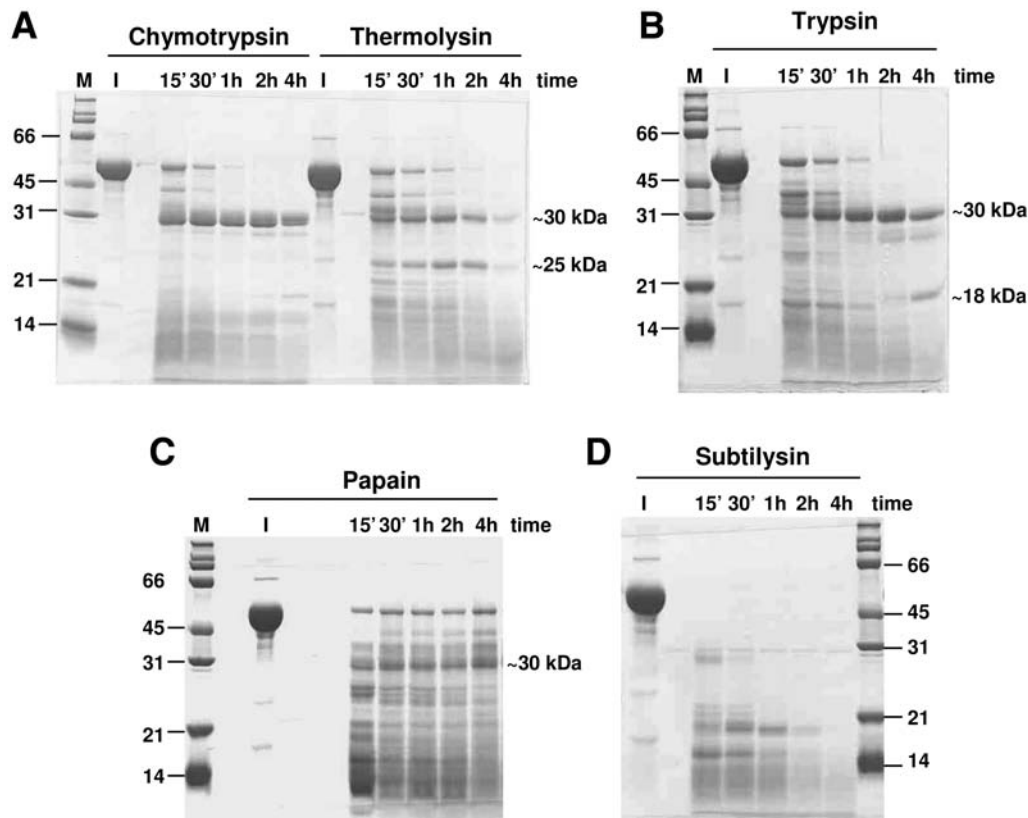


Fig. 1. SDS-PAGE analyses of AtINT6 partial proteolysis. The proteases used in the assays are indicated above each panel, and the incubation time is indicated above each lane. I, untreated AtINT6 protein. The molecular weight marker and the protease resistant products also are indicated.

29,623.96 Da and matches exactly the trypsin-resistant fragment identified by MALDI-TOF. The approx 30-kDa thermolysin product (encompassing residues 124–385) does not correspond exactly to the approx 30-kDa trypsin (encompassing residues 167–418), suggesting that the protease-resistant core fragment is located in the region between amino acids 167–175 and 385–418 (Fig. 2). Note that this region overlaps but does not match exactly the region identified by sequence alignment in the C-terminal half that contains approx 200 amino acids of the PCI domain and characterized by the higher sequence conservation (1,2). For AtINT6, the conserved C-terminal half includes amino acids 288–404 (Conserved Domain Database, www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=pfam01399). In conclusion, however, limited proteolysis analyses revealed that the resistant fragments include the PCI domain identified by sequence alignment with additional amino acids at the N terminus and possibly also at the C terminus.

Construction and Expression of Truncated Proteins Containing the PCI Domain

The finding that the trypsin and thermolysin proteolysis products of AtINT6 show similar molecular mass

but differ at the N- and C-terminal regions led us to construct vectors to express two truncated proteins with N and C terminus tentatively determined by using both the MS information and secondary structure prediction. Amino acid 124 corresponds to the N terminus of the thermolysin largest resistant product, and amino acid 418 corresponds to the C terminus of the trypsin largest fragment. The truncated protein AtINT6TR2 contains this region (Fig. 4D), although its C terminus was set at amino acid 415 instead of 418 to better fit the scores of secondary structure prediction (14), which indicate low scores for amino acids 417–418 to form α -helix (data not shown). Indeed, by removing these amino acids, we intended to generate a truncated protein without unstructured regions at the C terminus. In parallel, a second truncated protein, AtINT6TR1, containing amino acids 124–441 was constructed to obtain a protein with the native C terminus of AtINT6 (Fig. 4D), which we expected to positively affect its stability. Surprisingly, AtINT6TR1 was totally insoluble after expression in *E. coli* (Fig. 4A,B), suggesting that the 26 C-terminal amino acids negatively interfere with the folding of this truncated protein. Relatively high levels of soluble AtINT6TR2 were obtained at lower incuba-

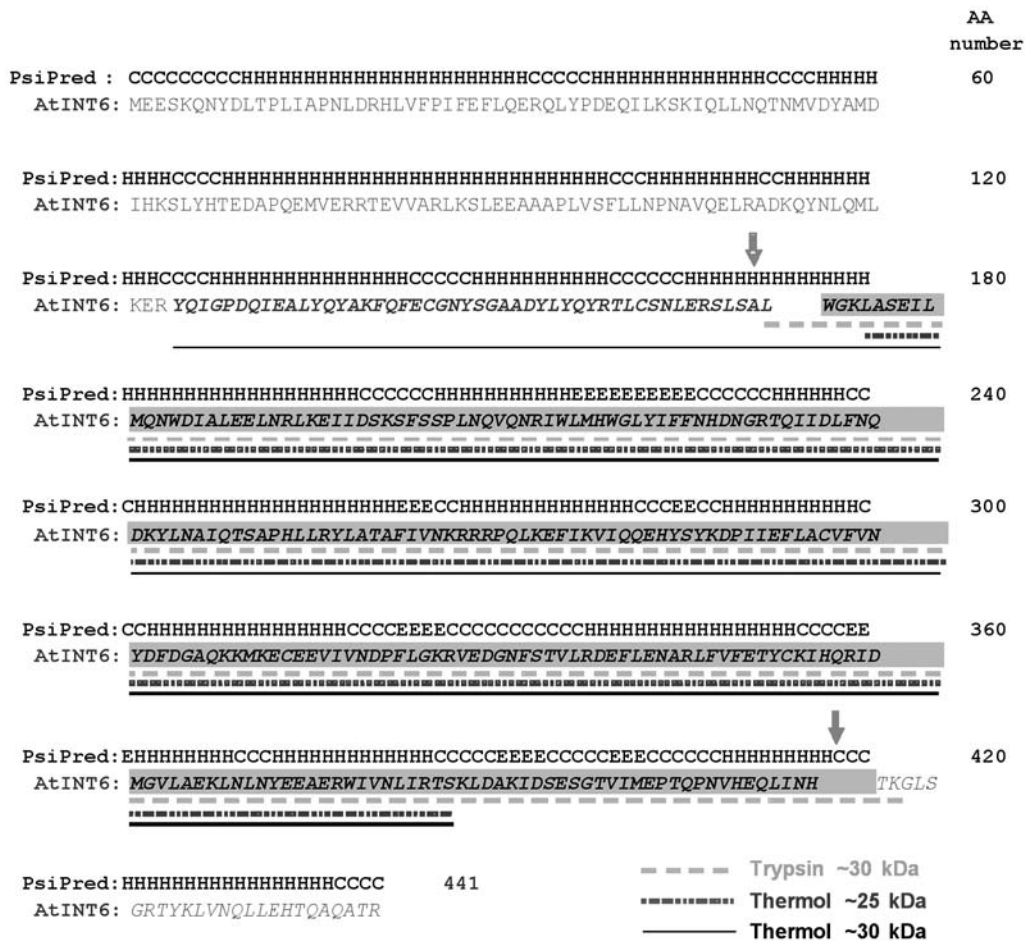


Fig. 2. AtINT6 amino acid sequence showing the protease-resistant fragments identified by MALDI-TOF fingerprinting and electrospray ionization MS. Lines underneath the amino acid sequence indicate the approx 30-kDa thermolysin-resistant fragment (solid line), the approx 25-kDa thermolysin-resistant fragment (dotted line), and the approx 30-kDa trypsin-resistant fragment (dashed line). Arrows indicate the fragment from residue 167 to 418 identified using electrospray ionization. The secondary structure (predicted using PSIPRED) (11) is shown above the amino acid sequence. H, α -helix; E, β -sheet; C, coil. The amino acid sequence corresponding to AtINT6TR1 is shown in italics, AtINT6TR2 is in bold, and AtINT6TR3 is marked with a gray box.

tion temperatures, such as 25°C (Fig. 4A), whereas incubation of *E. coli* cells above this temperature resulted in production of insoluble protein.

A third truncated form, AtINT6TR3, was constructed to adjust the N terminus more closely to the trypsin proteolytic product. This product contains amino acids 167–418 and was identified by both MALDI-TOF and electrospray ionization MS. The C terminus of AtINT6TR3 corresponds to the C terminus of AtINT6TR2, and the N terminus was set at amino acid 172 instead of 167 because, according to the secondary structure prediction using PSIPRED (14), amino acids 167–171 show a low score for α -helix formation (data not shown). Thus, as for the C terminus, we intended to remove all unstructured amino acids from the N terminus of AtINT6TR3 (Fig. 4D). As shown in Figure 4, this strategy was very effective to

generate a truncated protein showing high level of expression in *E. coli*. AtINT6TR3 also showed high solubility, being soluble even in cells incubated at higher temperatures, thus suggesting it to be more stable than AtINT6TR2. AtINT6TR3, made up of amino acids 172–415, contains the predicted PCI domain identified by bioinformatics and should serve as a model protein for the structural characterization of the PCI domain.

Conformational and Stability Analyses of AtINT6 and of Its Truncated Forms

The purified recombinant proteins (Fig. 4C) showed CD spectra with high helical content (Fig. 5), as predicted for the PCI domain and for the AtINT6 full-length protein (1,2). The CD spectra obtained for AtINT6 and AtINT6TR2 are very similar. Interestingly,

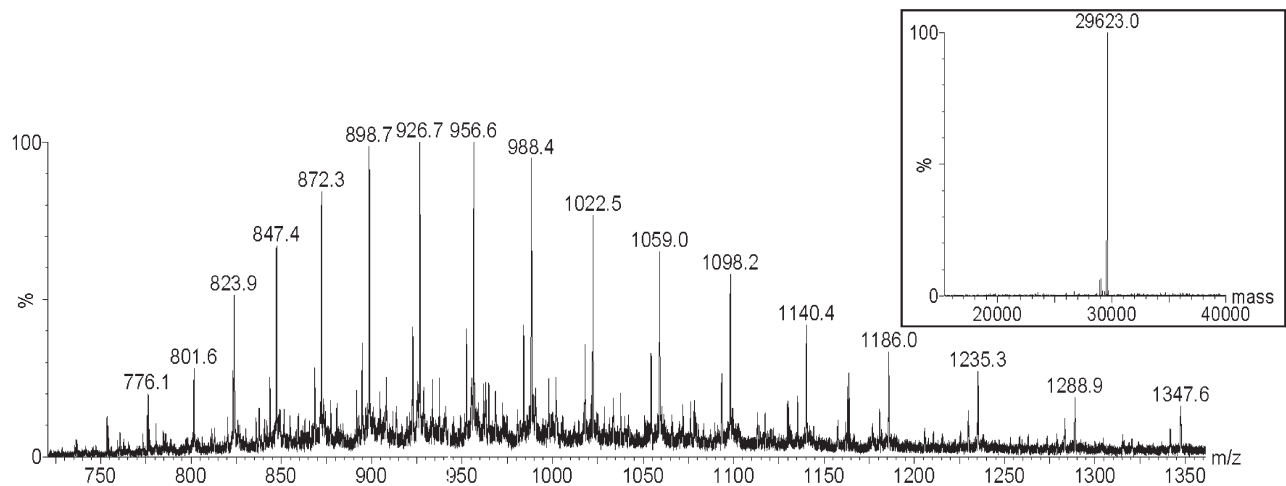


Fig. 3. Electrospray MS spectrum of AtINT6 trypsin-resistant domain. The inset shows the deconvoluted spectrum and the actual mass of the domain.

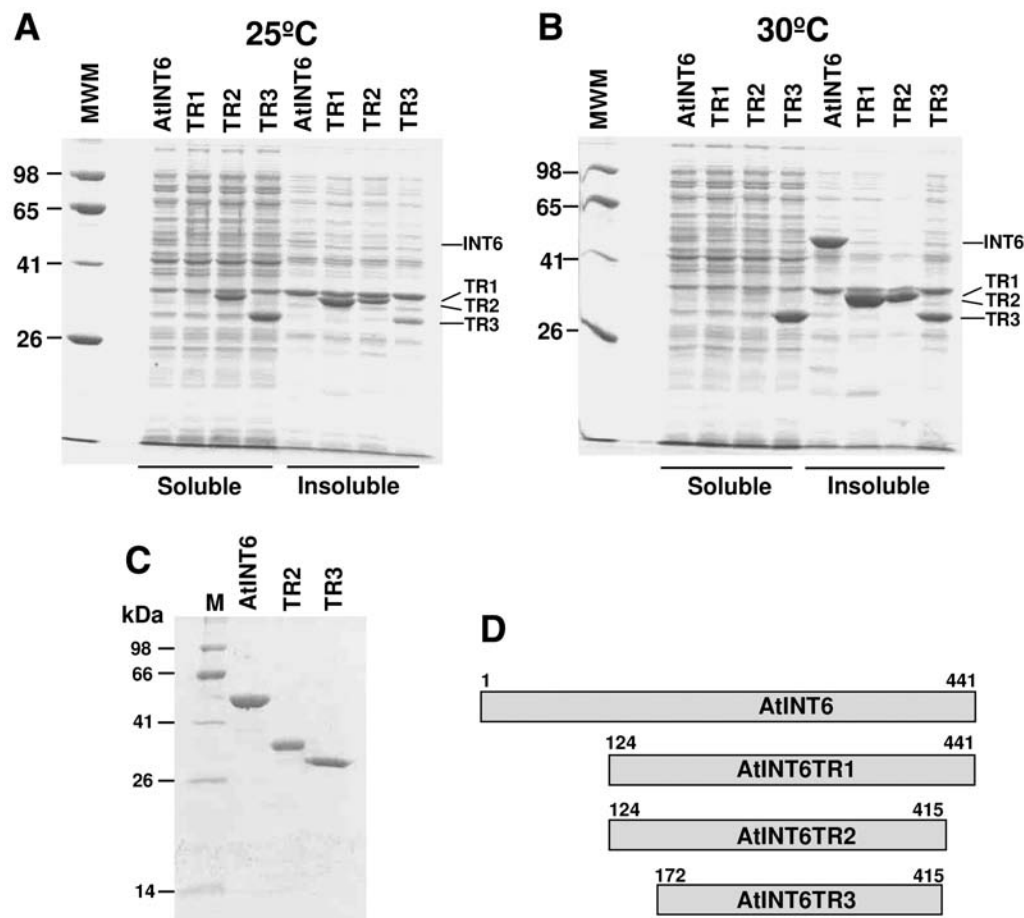


Fig. 4. Expression of the AtINT6 truncated forms in *E. coli*. (A and B) SDS-PAGE analyses of the recombinant proteins expressed in *E. coli* at 25 and 30°C, respectively. Soluble and insoluble fractions are indicated at the bottom of each panel. (C) SDS-PAGE gel showing the purified recombinant proteins. (D) Scheme representing the AtINT6 deletion mutants.

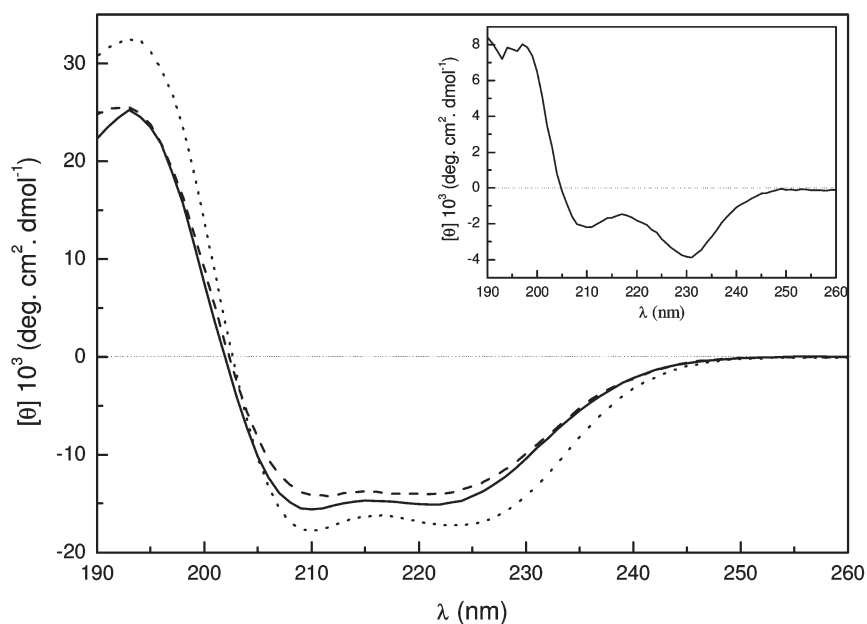


Fig. 5. Circular dichroism spectra of AtINT6 and of its truncated forms. Spectra of the full-length AtINT6 (solid line) and the truncated proteins AtINT6TR2 (dashed line) and AtINT6TR3 (dotted line). Inset, difference spectrum between AtINT6TR3 and AtINT6.

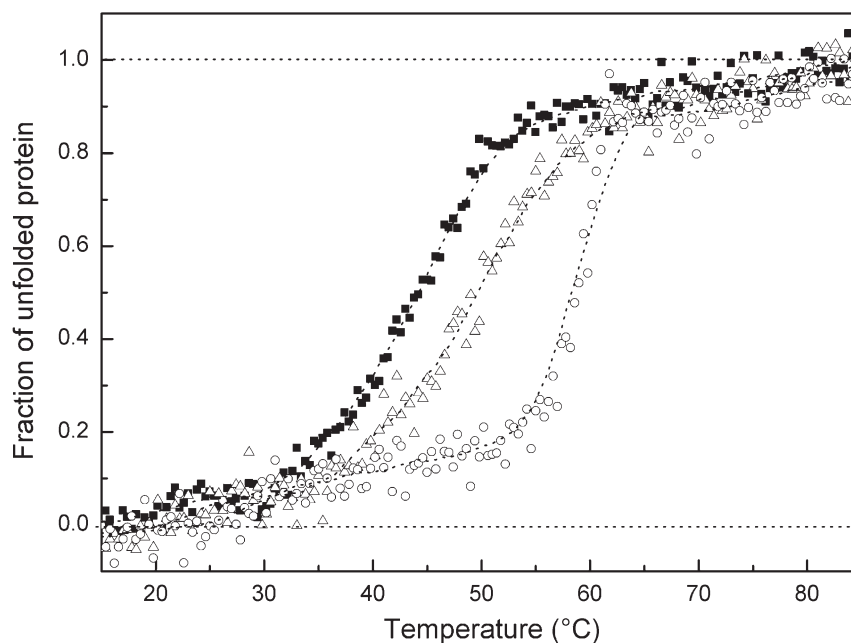


Fig. 6. Thermal denaturation of AtINT6 (■), AtINT6TR2 (△), and AtINT6TR3 (○) measured at 222 nm. The curves show the fraction of unfolded protein relative to increasing temperatures.

AtINT6TR3 exhibits a hyperoptical activity and the band at 220 nm shifted to 225 nm. Indeed, the AtINT6TR3–AtINT6 difference spectrum (Fig. 5, inset) reveals a negative band at 230 nm that can be assigned to the lower content of aromatic residues in AtINT6TR3

(28) compared with AtINT6 (46) and AtINT6TR2 (37). Indeed, aromatic residues show a positive contribution in this region of the CD spectrum (15). In addition, the positive and negative peaks at 195 and 208 nm, respectively, of the difference spectrum indicate that

AtINT6TR3 is characterized by a higher content of α -helix (Fig. 5, inset). However, the overlapping of the 222-nm contribution because of the α -helix with the aromatic contribution at 230 nm does not allow quantitative estimation.

The differences observed in the expression levels of the recombinant proteins indicate that folding and stability of AtINT6TR3 should be favored compared with AtINT6TR2 and full-length AtINT6. Thermal unfolding was followed by CD and showed that AtINT6 and AtINT6TR2 possess midpoint transition temperatures of 44 and 49°C (Fig. 6), respectively. AtINT6TR3, in contrast, exhibited a significantly higher midpoint transition temperature of 58°C, with a sharper transition from the folded to the unfolded state (Fig. 6), confirming the higher structure stability and a stronger cooperativity of the transition. After thermal denaturation at 90°C, refolding was found to be irreversible for the three proteins (data not shown). It is intriguing that the 26 amino acids of the C terminus have such a strong negative effect on the solubility of AtINT6TR1, and we have no explanation so far. The effect of deleting the sequence between amino acids 124 and 172 was striking, because AtINT6TR3 shows higher expression levels, solubility, and thermal stability than AtINT6TR2. In this case, however, we feel that these differences are due to the high content of hydrophobic residues of the 124–172 region that also includes two of the five cysteine residues of the protease resistant product (Fig. 2).

Characterizing the PCI domain structure is essential to understanding its function because the mouse INT6-truncated proteins lacking the PCI domain show cell-transforming activity (4,5). Furthermore, INT6 protein–protein interaction activity is mediated by the PCI domain region (2,12), and no clues are available to explain how the PCI domain can specifically recognize different binding partners. In conclusion, the combination of limited proteolysis, MS, and CD spectroscopy used in this work allowed us to identify a truncated protein showing high level of expression, and, more importantly, high solubility and stability. This truncated protein, AtINT6TR3, contains the PCI domain identified by bioinformatics with some additional amino acids at the N- and C-termini and has the potential to serve as a model protein for the three-dimensional structure determination of the PCI domain.

ACKNOWLEDGMENTS

We are grateful to Celso Eduardo Benedetti for providing *A. thaliana* mRNA; to Markus B. Smolka and José Camillo Novello for the peptide fingerprinting analysis; and to Patrícia R. Marques, Tereza C. Lima Silva,

Adriana C. Alves, Luciana Rodrigues Camillo, and Zildene Gonçalves Correa for technical support.

REFERENCES

- Hofmann, K. and Bucher, P. (1998) The PCI domain: a common theme in three multiprotein complexes. *Trends Biochem. Sci.* **23**, 204–205.
- Kim, T., Hofmann, K., von Arnim, A. G., and Chamovitz, D. A. (2001) PCI complexes: pretty complex interactions in diverse signaling pathways. *Trends Plant Sci.* **6**, 379–386.
- Marchetti, A., Buttitta, F., Miyazaki, S., Gallahan, D., Smith, G. H., and Callahan, R. (1995) Int-6, a highly conserved, widely expressed gene, is mutated by mouse mammary tumor virus in mammary preneoplasia. *J. Virol.* **69**, 1932–1938.
- Rasmussen, S. B., Kordon, E., Callahan, R., and Smith, G. H. (2001) Evidence for the transforming activity of a truncated Int6 gene, in vitro. *Oncogene* **20**, 5291–5301.
- Mayeur, G. L. and Hershey, J. W. (2002) Malignant transformation by the eukaryotic translation initiation factor 3 subunit p48 (eIF3e). *FEBS Lett.* **514**, 49–54.
- Asano, K., Merrick, W. C., and Hershey, J. W. (1997) The translation initiation factor eIF3-p48 subunit is encoded by int-6, a site of frequent integration by the mouse mammary tumor virus genome. *J. Biol. Chem.* **272**, 23,477–23,480.
- Guo, J. and Sen, G. C. (2000) Characterization of the interaction between the interferon-induced protein P56 and the Int6 protein encoded by a locus of insertion of the mouse mammary tumor virus. *J. Virol.* **74**, 1892–1899.
- Guo, J., Hui, D. J., Merrick, W. C., and Sen, G. C. (2000) A new pathway of translational regulation mediated by eukaryotic initiation factor 3. *EMBO J.* **19**, 6891–6899.
- Desbois, C., Rousset, R., Bantignies, F., and Jalinot, P. (1996) Exclusion of Int-6 from PML nuclear bodies by binding to the HTLV-I Tax oncoprotein. *Science* **273**, 951–953.
- Moris-Desbois, C., Bochar, V., Reynaud, C., and Jalinot P. (1999) Interaction between the Ret finger protein and the int-6 gene product and co-localization into nuclear bodies. *J. Cell Sci.* **112**, 3331–3342.
- Watkins S. J. and Norbury, C. J. (2004) Cell cycle-related variation in subcellular localization of eIF3e/INT6 in human fibroblasts. *Cell Prolif.* **37**, 149–160.
- Yahalom, A., Kim, T. H., Winter, E., Karniol, B., von Arnim, A. G., and Chamovitz, D. A. (2001) Arabidopsis eIF3e (INT-6) associates with both eIF3c and the COP9 signalosome subunit CSN7. *J. Biol. Chem.* **276**, 334–340.
- Ausubel F. M., Brent R, Kingston R. (1998) *Current Protocols in Molecular Biology*, New York: Wiley.
- McGuffin L. J., Bryson K, and Jones D. T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405.
- Sreerama, N., Manning, M. C., Poweres, M. E., Zhang, J.-X., Goldenberg, D. P., and Woody, R. W. (1999) Tyrosine, phenylalanine, and disulfide contributions to the circular dichroism of proteins: circular dichroism spectra of wild-type and mutant bovine pancreatic trypsin inhibitor. *Biochemistry* **38**, 10,814–10,822.

3.3-Artigo II

Spectroscopic characterization of the tumor antigen NY-REN-21 and identification of heterodimer formation with SCAND1

Flávia R. G. Carneiro, Tereza C. L. Silva, Adriana C. Alves, Thais Haline-Vaz, Fábio C. Gozzo and Nilson I. T. Zanchin

Spectroscopic characterization of the tumor antigen NY-REN-21 and identification of heterodimer formation with SCAND1

Flávia R.G. Carneiro, Tereza C.L. Silva, Adriana C. Alves, Thais Haline-Vaz,
Fabio C. Gozzo, Nilson I.T. Zanchin *

*Center for Structural Molecular Biology, Brazilian Synchrotron Light Laboratory, LNLS, Rua Giuseppe Maximo Scolfaro 10000,
P.O. Box 6192, CEP 13084-971, Campinas SP, Brazil*

Received 21 February 2006
Available online 3 March 2006

Abstract

Human NY-REN-21 is a C₂H₂ type multi-finger protein, with a SCAN domain in the N-terminal region and a predicted coil central region. It represents a putative ortholog of mouse ZFP38, a transcriptional factor that recognizes a bipartite DNA motif and is unable to form homodimers. As shown in this work, NY-REN-21 contains a SCAN domain able to form homodimers and a central region that behaves as an intrinsically disordered protein. The SCAN domain is found in 71 human proteins and its ability to form homo- and heterodimers widens the number of genes that are regulated by this group of transcription factors. NY-REN-21 interaction with SCAND1 was identified using the yeast two-hybrid system and confirmed using recombinant proteins. SCAND1 is a truncated SCAN box protein, lacking the zinc finger region and the NY-REN-21/SCAND1 heterodimer is asymmetric concerning the DNA binding region. This result indicates that NY-REN-21 can function either as a homodimer or as a heterodimer with SCAND1.

© 2006 Elsevier Inc. All rights reserved.

Keywords: SCAN domain protein family; Intrinsically disordered protein; Zinc finger protein; Circular dichroism; Mass spectrometry; Yeast two-hybrid system

NY-REN-21 was identified in a SEREX screen (*serological analysis of recombinant cDNA expression libraries*) as an autologous antibody in patients with renal-cell carcinoma [1]. NY-REN-21 belongs to the multi-finger SCAN box protein family, containing the SCAN domain [2] in the N-terminal region and seven C₂H₂ type zinc fingers in the C-terminal region. NY-REN-21 also contains a 150 residue central region predicted to be intrinsically disordered under physiological conditions. Zinc finger transcriptional factors can exert either positive or negative regulation of gene targets [3]. Diversity in the modular multi-finger protein family is determined by the conserved domains found outside the zinc finger region. These domains include BTB/POZ [4], KRAB [5], and SCAN [2]. They usually control

selective association of the transcription factors with other cellular factors essential for regulation, subcellular localization, DNA binding or transcription. The SCAN domain is restricted to vertebrates, acting as a mediator of both homo and heterodimerization [6–10]. It is an 82–88 amino acid highly conserved domain [7,8,10,11], forming a structurally independent module with 5 distinct α -helices in each monomer [12].

NY-REN-21 represents a putative ortholog of mouse ZFP38, also named CTfin51, Zipro1, and Ru49, sharing 82% amino acid identity. An important difference is observed in the N-terminal region where NY-REN-21 lacks 77 residues of the trans-activation region relative to ZFP38 (Fig. 1). ZFP38 functions as a transcriptional factor, containing three sequence repeats that show trans-activation activity [13]. It binds to a bipartite motif (5'AGTAC3') showing higher affinity to motifs separated by 2–7 nucleotides [14]. Studies using transgenic mice

* Corresponding author. Fax: +55 19 3512 1004.
E-mail address: zanchin@lnls.br (N.I.T. Zanchin).

Fig. 1. Amino acid sequence alignment of the human NY-REN-21 and mouse ZFP38 proteins. The conserved domains are indicated for NY-REN-21 as follows: italic-bold, acidic region (residues 17–33); underlined, SCAN domain (residues 36–131); gray box, disordered central region (residues 128–276); bold, zinc finger region (residues 275–473).

Up to now there are no experimental data available on NY-REN-21 structure and little is known about its function on transcriptional control. In this work, we present data on the spectroscopic characterization of NY-REN-21, confirming the helical structure of the SCAN domain and the intrinsically disordered nature of the central region. Unlike ZFP38, which apparently does not form homodimers [8], the SCAN domain of NY-REN-21 is able to form homodimers. In addition, NY-REN-21 forms heterodimers with SCAND1, which is a SCAN family protein that lacks the zinc finger region [7]. Heterodimerization between different members of the SCAN protein family probably allows them to control expression of a larger number of genes and the identification of the NY-REN-21-SCAND1 interaction may help understand the mechanism of action of the multi-finger transcription factors.

Construction of expression vectors. The NY-REN-21 cDNA (Accession No. [NM_145914](#)) was amplified by PCR from a human fetal brain library (Clontech HL4028AH) using primers ONZ111 and ONZ112 (oligonucleotides are listed in [Table 1](#)) and inserted into the *Nde*I and *Eco*RI sites of pET28a (Novagen) generating vector pET-NYREN21. pET-SCAN

pGAD-C2 [16] (activation domain fusions) and pTL1 (DNA binding domain fusions) were used for yeast two-hybrid assays. pTL1 is a pBTM116 [17] derivative containing the *Escherichia coli* kanamycin marker. pTL1-SCAN and pGAD-SCAN contain the SCAN domain amplified using primers ONZ288 and ONZ289, and inserted into the *Bam*HI and *Eco*RI sites of pTL1 and pGAD-C2, respectively. pACT-SCAND1 contains the cDNA of SCAND1 isolated from the human fetal brain cDNA library. pET3-SCAND1 was constructed by inserting a *Bg*III HA-SCAND1 fragment isolated from pACT-SCAND1 into the *Bam*HI site of pET3d (Novagen).

Protein expression and purification. *Escherichia coli* BL21(DE3)*slyD*⁻ (a gift from Ryland Young, Texas A&M University) was used to express the recombinant proteins and the cultures were induced with 0.5 mM IPTG at an OD₆₀₀ ~0.8. The induction time was 6 h at 20 °C for NY-REN-21; 4 h at 37 °C for the SCAN domain and GST-DCR; 4 h at both 25 °C and 37 °C for GST-ZFR and His-ZFR. Induction of GST-ZFR was also performed with 10 mM lactose for 16 h at 20 °C. Extracts were

Table 1
Oligonucleotides used in PCRs

Number	Sequence	Sites
ONZ111	5'TTACCATATGACCAAGGTACTAGGCATGGC 3'	<i>NdeI</i>
ONZ112	5' TGAATTCCTACGGTGCTTCTCCCTCTCCAG 3'	<i>EcoRI</i>
ONZ288	5' GTACGAATTCAAGTACCTTCTAGCCTG 3'	<i>EcoRI</i>
ONZ289	5' ACTGGATCCTCATCCTGGCTCATCCAGTTC 3'	<i>BamHI</i>
ONZ290	5' ACGGATCCCTAATGAAGTACCTTCTAGCCTG 3'	<i>BamHI</i>
ONZ291	5'GACTGAATTCTCATCCTGGCTCATCCAGTTC 3'	<i>EcoRI</i>
ONZ321	5' CTTAGAATTCAGACGTTATATATGTGCTG 3'	<i>EcoRI</i>
ONZ322	5' GTACCTCGAGTTACGGTGCTTCTCCCTCTC 3'	<i>XhoI</i>
ONZ323	5' GACTGAATTCGATGAGCCAGGACACCAAGG 3'	<i>EcoRI</i>
ONZ324	5'GATACTCGAGTTAACGTCTCTCTCTGGGG 3'	<i>XhoI</i>

prepared in the appropriated buffers and cleared by centrifugation. Purification of histidine- and GST-tagged proteins was performed on standard affinity chromatography on metal-chelating and glutathione–Sepharose columns, respectively. NY-REN-21 extracts were prepared in buffer A (20 mM Hepes, pH 7.2, 100 mM NaCl, 20% glycerol (v/v), 5 mM β -mercapto-ethanol, 10 μ M ZnSO₄, and 0.5 mM PMSF) and treated with DNase I (2.5 μ g/mL) for 40 min at room temperature prior to affinity chromatography. Affinity-purified NY-REN-21 was fractionated on a 1 mL heparin–Sepharose column equilibrated in buffer B (buffer A + 300 mM NaCl) and eluted with a 15-mL gradient of buffer C (buffer A + 1.5 M NaCl). SCAN domain extracts were prepared in buffer D (50 mM sodium phosphate, pH 7.0, 100 mM NaCl, and 5% glycerol) and purified using a nickel-chelating column. Following removal of the histidine-tag by TEV protease digestion, the domain was further purified using a 1 mL MonoQ column with buffer E as binding buffer (25 mM sodium phosphate, pH 7.0, 5% glycerol, 10 mM NaCl, and 1 mM DTT) and a gradient of buffer F (buffer E + 1 M NaCl) for elution. GST-DCR extracts were prepared in phosphate-buffered saline containing 5 mM β -mercapto-ethanol and 0.5 mM PMSF, and GST-DCR was affinity-purified using glutathione–Sepharose beads (GE Healthcare) according to the supplier's instructions. GST-DCR was cleaved using the TEV protease and GST removed by incubating the suspension extensively with glutathione–Sepharose beads. GST-ZFR affinity chromatography on glutathione–Sepharose beads was performed as described for SCAN-DCR.

TEV protease digestion and limited proteolysis assays. TEV protease digestions were performed in buffer containing 50 mM Tris–HCl, pH 8.0, 5 mM EDTA, and 1 mM DTT for 16 h at 25 °C (TEV protease/fusion protein ratio 1:20). For limited proteolysis, purified DCR and SCAN domain were subjected to digestion with trypsin (1%, w/w) at room temperature in buffer containing 50 mM sodium phosphate, pH 7.2, 100 mM NaCl, 5% glycerol, and 0.5 mM DTT. Samples were collected at various times and the reactions were stopped by adding Tricine–SDS–PAGE sample buffer containing 1 mM PMSF and heated at 95 °C for 5 min. The proteolysis products were analyzed on Tricine–SDS–PAGE gels.

Circular dichroism and fluorescence spectroscopy. Circular dichroism (CD) spectra were acquired on a Jasco-810 spectropolarimeter at 20 °C using a 50 nm/min scanning rate and a spectral bandwidth of 0.5 nm. Ellipticity is reported as the mean residual ellipticity [θ] (deg cm² dmol^{−1}). CD spectra of NY-REN-21 were acquired at 82 μ M using a 0.01 mm path length cell over the range of 190–260 nm in buffer containing 20 mM Hepes, pH 7.2, 900 mM NaCl, 20% glycerol (v/v), 5 mM β -mercapto-ethanol, 10 μ M ZnSO₄, and 0.5 mM PMSF. EDTA-treated samples were incubated in the same buffer with 2 mM EDTA for 30 min prior spectrum acquisition. The CD spectra of DCR and the SCAN domain at 10 μ M in 10 mM sodium phosphate buffer, pH 7.2, 20 mM NaCl, and 1 mM DTT were acquired using a 0.1 mm path length cell over the range of 190–260 nm. The SCAN domain was subjected to thermal unfolding at 5 μ M and 10 μ M from 10 °C to 95 °C. Ellipticity at 222 nm was measured using 0.5 °C intervals and spectra were collected at 5 °C intervals over the λ range of 190–260 nm. Midpoint transition temperatures were calculated as the center of the Gaussian fit of the first derivative of the denaturation curves. Refolding assays started at 95 °C and the temperature lowered to 20 °C. DCR intrinsic fluorescence was measured at 10 μ M in 10 mM sodium phosphate buffer, pH

7.2, 20 mM NaCl, and 1 mM DTT using an ISSK2 fluorimeter. Excitation was at 280 nm and emission at 340 nm was recorded from 300 to 400 nm.

Mass spectrometry. The SCAN domain (6 mg/mL) in 100 mM ammonium bicarbonate buffer, pH 8.0, was analyzed by direct infusion electrospray ionization on a Q-ToF Ultima API mass spectrometer (Waters/Micromass) with a nanoflow interface. The instrument settings were 3 kV for the spray voltage, 100 V for the cone voltage, cone gas at 30 L/h, and source temperature of 100 °C. The final spectra were processed using the MaxEnt 1 program (Waters/Micromass). For mass spectrometry analysis, the recombinant DCR was dialyzed against 100 mM ammonium bicarbonate buffer, pH 8.0, containing 5 mM β -mercapto-ethanol, dried, and suspended in water/acetonitrile (1:1, v/v) containing 0.1% TFA. Data were acquired and analyzed as described for the SCAN domain.

Yeast two-hybrid assays. The yeast two-hybrid system was used to test homodimerization of the NY-REN-21 SCAN domain. The assay was performed in strain L40 [18] co-transformed with vectors pTL1-SCAN and pGAD-SCAN. The test and control strains were incubated in YNB medium supplemented with adenine and either histidine or 15 mM 3-amino-1,2,4-triazole (3-AT). Filter assays to test β -galactosidase expression were performed as described previously [18]. Strain L40 transformed with pBTM-Nip7 and pACT-Nop8 was used as a positive control [19]. The combinations pBTM-Nop8/pGAD-SCAN [19], pTL1/pGAD-SCAN, and pTL1-SCAN/pGAD2 co-transformed into L40 served as negative controls. A large-scale screen was performed to identify the NY-REN-21 SCAN domain-interacting partners. For this purpose, strain L40 containing pTL1-SCAN was transformed with a human fetal brain cDNA library fused to the activation domain of pACT2 (Clontech). Positive clones were selected on YNB medium supplemented with adenine and 7 mM 3-AT, followed by β -galactosidase filter assays [18].

Protein interaction assays. The proteins were co-expressed in *E. coli* BL21(DE3)*slyD*[−] using the combinations pET3-SCAN1/pET-NYREN21, pET3-SCAN1/pET-SCAN, and pET3-SCAN1/pET-Nip7 (pET-Nip7, unpublished results). Extracts were prepared in affinity buffer (100 mM sodium phosphate, pH 7.2, 100 mM NaCl, 5% glycerol, and 0.5 mM PMSF) and incubated with 25 μ L Ni–NTA agarose beads (Qiagen) for 45 min at 4 °C. Subsequently, the agarose beads were washed with 5 mL of wash buffer (affinity buffer + 200 mM NaCl) and the bound proteins eluted in 50 μ L of wash buffer containing 200 mM imidazole. The bound proteins were analyzed by SDS–PAGE and SCAN1 was identified by immunoblotting using an anti-HA antiserum (Santa Cruz Biotechnology).

Results and discussion

Expression and CD analysis of NY-REN-21

NY-REN-21 is a modular protein with a SCAN domain at the N-terminal region, a predicted disordered central region (DCR), and seven C₂H₂-type zinc fingers (ZFR) at the C-terminal region (Fig. 1). The central and the zinc finger regions confer high instability to the protein. Nevertheless,

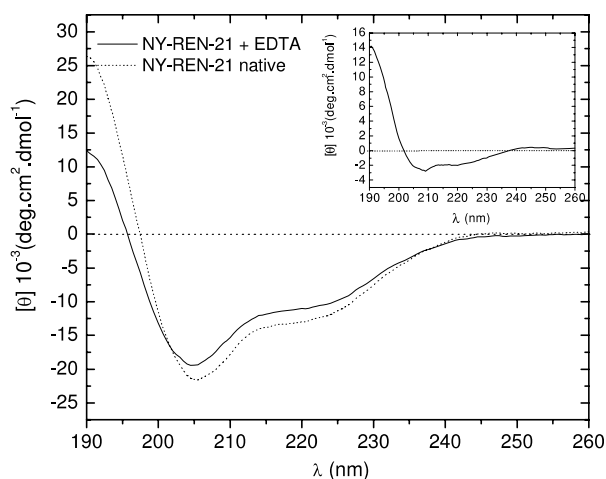


Fig. 2. Circular dichroism analysis of NY-REN-21. The dotted line represents the untreated protein and the solid line the sample treated with 2.0 mM EDTA for 30 min prior to spectrum acquisition. Inset: difference spectrum between NY-REN-21 native and NY-REN-21 treated with EDTA.

high levels of NY-REN-21 were obtained by inducing *E. coli* cells at 20 °C during 6 h. The recombinant protein was purified by affinity followed by a chromatography on heparin and was relatively stable in the heparin elution buffer containing 20 mM Hepes, pH 7.2, ~0.9 M NaCl, 20% glycerol (v/v), 5 mM β -mercapto-ethanol, 10 μ M ZnSO₄, and 0.5 mM PMSF. Dialysis against lower salt buffers led NY-REN-21 to aggregate and precipitate (data not shown). The CD spectrum of NY-REN-21 revealed a partially unfolded protein with a deeper minimum at 205 nm that must represent the contribution of the central coil region and possibly of partially unfolded zinc finger domains (Fig. 2). The C₂H₂ zinc finger is a structurally independent domain showing a typical β – β – α compact fold in which the zinc atom is bound by two cysteine and two histidine residues [20]. The zinc finger is prone to oxidation being highly unstable particularly when not bound to the cognate DNA [21]. In addition, chelating of the zinc ion by EDTA causes loss of the α -helix content of zinc finger proteins [22]. As expected, the CD spectrum of NY-REN-21 treated with 2 mM EDTA showed a moderate reduction of signal in the λ 200–230 nm range and a sharper reduction in the 190–200 nm range (Fig. 2). This loss of structure upon EDTA treatment indicates that the zinc finger region was at least partially folded. In order to analyze the contribution of each NY-REN-21 region to the final structure of the protein and to obtain data about their respective functions, we constructed truncated proteins and their characterization is described in the following sections.

Analysis of the disordered central region and the zinc finger region

Sequence analysis algorithms [23,24] suggested that the NY-REN-21 region comprising amino acids 128–276 falls

into the group of intrinsically disordered proteins. GST-DCR showed an aberrant electrophoretic mobility, migrating with a molecular weight on SDS–PAGE gels similar to GST-ZFR, which is 50 amino acids longer (Figs. 1 and 3A). Mass spectrometry was used to determine whether the correct DCR was produced. Its predicted molecular mass is 16967.5 Da and the mass spectrum of DCR cleaved from GST by TEV protease revealed a product of 16965.25 Da, which is quite close to the predicted molecular mass, and two additional products of 17041.25 and 17119.25 Da (Fig. 3B). The additional products correspond, respectively, to DCR containing one and two β -mercapto-ethanol molecules bound to its cysteine residues. Since GST-DCR was purified in the presence of β -mercapto-ethanol, we concluded that DCR was correctly expressed in *E. coli*. Its aberrant electrophoretic mobility may be due to a higher resistance of the unfolded protein against the polyacrylamide gel. This type of aberrant mobility has been described for *Saccharomyces cerevisiae* Gir2, which is an intrinsically disordered protein [25].

The techniques most employed to demonstrate the intrinsically unstructured nature of proteins are proteolysis, CD, and fluorescence spectroscopy, although there has been an increasing use of nuclear magnetic resonance and small angle X-ray scattering [26,27]. In order to determine whether DCR was a natively unstructured protein, we used a proteolysis assay and CD and fluorescence spectroscopy. DCR was fully digested following 1 h of incubation with trypsin, whereas most of the SCAN domain, used as a control, remained undigested (Figs. 3C and D). This result was an important indication that DCR is unstructured. CD and fluorescence analyses supported this finding. The CD spectrum of DCR showed a minimum at 199 nm which is typical of disordered structure (Fig. 4A). DCR contains two tryptophan residues and its fluorescence emission spectrum shows a maximum near 355 nm (Fig. 4B) which is consistent with the fluorescence emission of free or water-exposed tryptophan residues of unfolded protein chains. The actual function of the DCR of NY-REN-21 is not known but it is speculated that the flexible region of these modular proteins is important for transcriptional complex assembly [28] or can act as a spacer, regulating the distance between adjacent domains [29].

NY-REN-21 is predicted to contain seven C₂H₂ zinc fingers (ZFR) in the C-terminal region (amino acids 275–473). The histidine-tagged ZFR was totally insoluble but the GST-ZFR could be purified from *E. coli* extracts, although its characterization was troublesome (Fig. 3A). A huge amount of DNA co-purified with ZFR even using the strategy that was successful to obtain DNA-free NY-REN-21 (data not shown). DNA affects protein quantitation and TEV protease cleavage was unsuccessful due to the quick degradation of the ZFR moiety of the fusion protein. Further characterization of the ZFR was not possible.

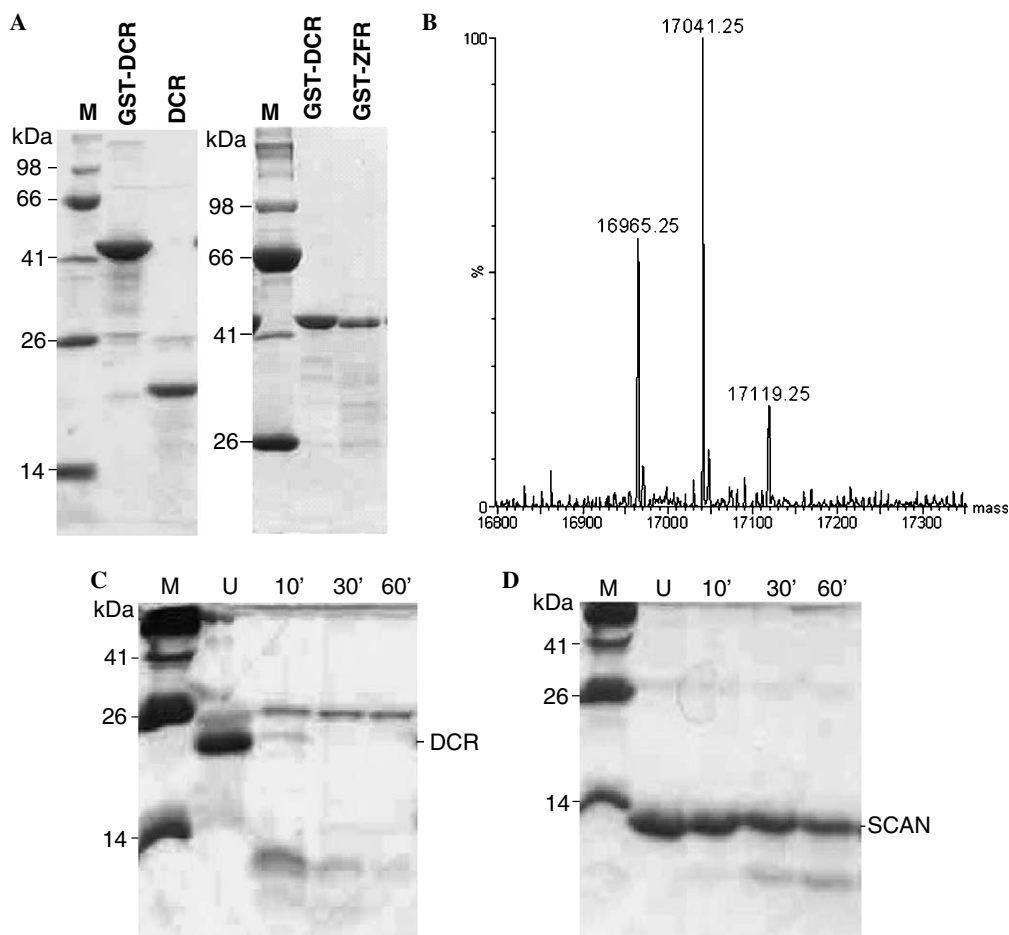


Fig. 3. Purification and analysis of the disordered central region (DCR). (A) Coomassie-stained SDS-PAGE gel showing purified DCR (left panel) and comparison of the electrophoretic mobility of GST-DCR and GST-ZFR (right panel). GST-DCR and GST-ZFR: respective purified fusion proteins; DCR: purified disordered central region following cleavage by TEV protease. (B) Mass spectrometry analysis of DCR. The deconvoluted spectra revealed products of 16965.25 Da corresponding to DCR (predicted molecular mass 16967.5 Da) and two additional products (17041.25 and 17119.25 Da) corresponding to DCR containing one and two β -mercapto-ethanol molecules, respectively. (C,D) Coomassie-stained SDS-PAGE gels showing trypsin proteolysis of DCR (C) and the SCAN domain (D), which was used as a control. The incubation time is indicated above each lane. U, Untreated protein. M, Molecular weight standards.

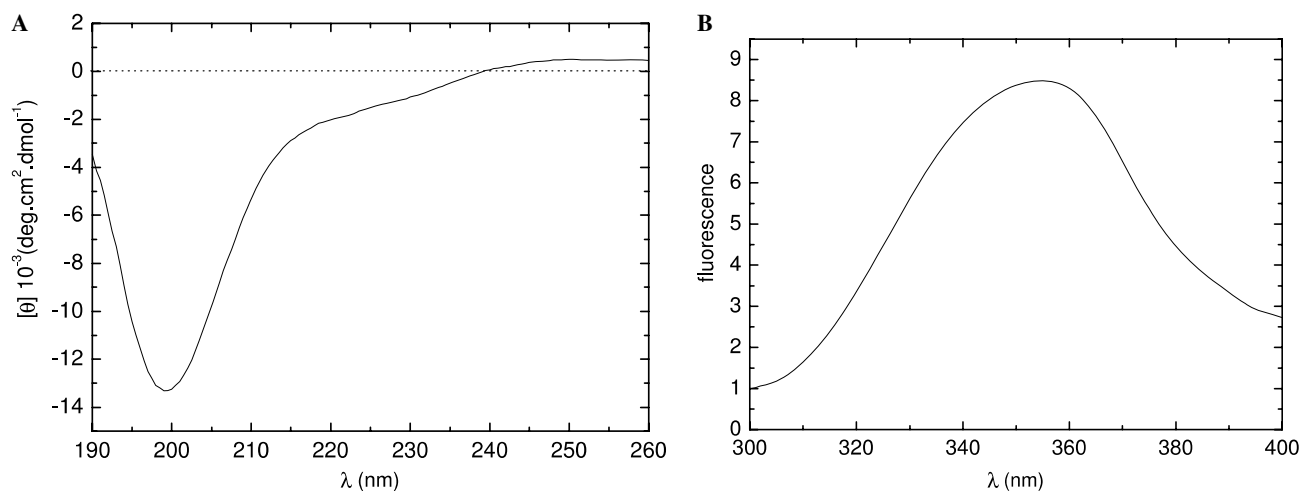


Fig. 4. Spectroscopic analysis of the disordered central region (DCR). (A) Circular dichroism spectrum of DCR showing a minimum at 199 nm. (B) Fluorescence analysis of DCR. The excitation and emission wavelengths used were 280 and 340 nm, respectively. The maximum fluorescence emission at 355 nm shows that the tryptophans are exposed to the solvent.

Secondary structure and oligomerization analysis of the SCAN domain

The SCAN domain showed a typical α -helical CD spectrum and could be partially refolded following thermal denaturation (Fig. 5A). The SCAN domain of ZNF174 was reported to show a concentration-dependent midpoint transition temperature as a consequence of the dimeric structure [10]. An unfolding behavior consistent with a dimer was also observed for the NY-REN-21 SCAN domain, which showed a midpoint transition temperature of 60 °C at 5 μ M and 62.8 °C at 10 μ M (Fig. 5B). Oligomerization of SCAN domains has been widely documented, forming either homo or heterodimers with other SCAN-containing proteins [6–10]. In the mammalian two-hybrid system, mouse ZFP38 did not form homodimers and interacted only with SRE-ZBP [8]. NY-REN-21

lacks 77 amino acids of the acidic region as compared to ZFP38 (Fig. 1) but this difference should not interfere with the oligomerization mediated by the SCAN domain. In order to determine whether the SCAN domain of NY-REN-21 was able to form homodimers, experiments were performed using the yeast two-hybrid system and mass spectrometry. The yeast two-hybrid assay was performed in strain L40 carrying the SCAN domain (amino acids 36–131) fused both to the *lexA* DNA binding domain and to the *GAL4* activation domain. This strain contains *HIS3* and *lacZ* as two-hybrid reporters, both of which were expressed at high levels, indicating that the SCAN domain formed homodimers with high affinity (Fig. 6). Parallel controls were performed with L40 transformed with vectors pBTM-Nop8/pGAD-SCAN, pTL1/pGAD-SCAN, and pTL-SCAN/pGAD-C2, which did not activate the two-hybrid markers (Fig. 6). Electrospray ionization analysis of the native SCAN domain in ammonium bicarbonate buffer identified a monomer as a minor product (14896 Da) and a dimer as a major product (29793 Da, Fig. 6C). The SCAN domain contains a cysteine residue and its oxidation could form a dimer linked by a disulfide bond. To exclude this possibility, a new analysis was performed under denaturing conditions in buffer containing 50% acetonitrile, 0.1% formic acid, without reducing the cysteine residues. The molecular mass obtained (14896 Da) corresponded to the monomer (Fig. 6D), indicating that the dimer was formed by specific, non-covalent interactions. Size-exclusion chromatography also showed a retention time consistent with the size of a dimer (data not shown). It is quite intriguing that no homodimer was described for mouse ZFP38 [8] since the two proteins share 97% amino acid similarity in the SCAN domain (Fig. 1).

Identification of the SCAN domain/NY-REN-21 interaction with SCAND1

The ability of this particular SCAN domain to interact with other SCAN family proteins was investigated by taking advantage of the yeast two-hybrid system. Five positive clones containing the cDNA encoding SCAND1 (also named SDP1 and RAZ1) were isolated from a human fetal brain library. The assay for the *HIS3* reporter was performed in medium containing high 3-AT concentration (up to 30 mM) which is a strong indication of positive interaction (Fig. 7A). Positive interaction was also observed for the *lacZ* reporter gene (Fig. 7B). Direct interaction of NY-REN-21 with SCAND1 was assayed using recombinant proteins. HA-tagged SCAND1 was co-expressed in *E. coli* with histidine-tagged NY-REN-21 or with histidine-tagged SCAN domain. A parallel control assay was carried out with HA-tagged SCAND1 co-expressed with histidine-tagged Nip7p, an unrelated protein. The histidine-tagged proteins were affinity-purified on nickel columns and SCAND1 copurified with both full-length NY-REN-21 and the SCAN domains of NY-REN-21 (Fig. 7C), confirming the heterodimer formation between these two proteins.

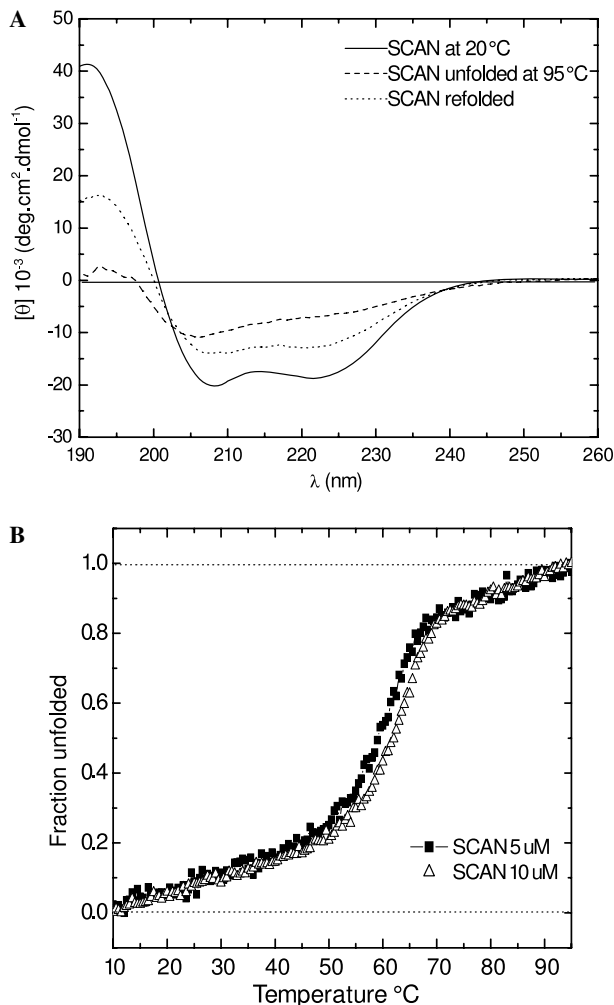


Fig. 5. Analysis of the recombinant SCAN domain. (A) CD spectra of the SCAN domain (10 μ M). Solid line, spectrum acquired at 20 °C; dashed line, spectrum acquired at 95 °C; dotted line, spectrum of refolded SCAN domain following thermal denaturation at 95 °C and refolding to 20 °C. (B) Thermal unfolding analysis of the SCAN domain. The molar ellipticity at 222 nm was determined at 5 μ M (■) and 10 μ M (Δ) in the range from 10 °C to 95 °C and plotted as a ratio of folded versus unfolded protein.

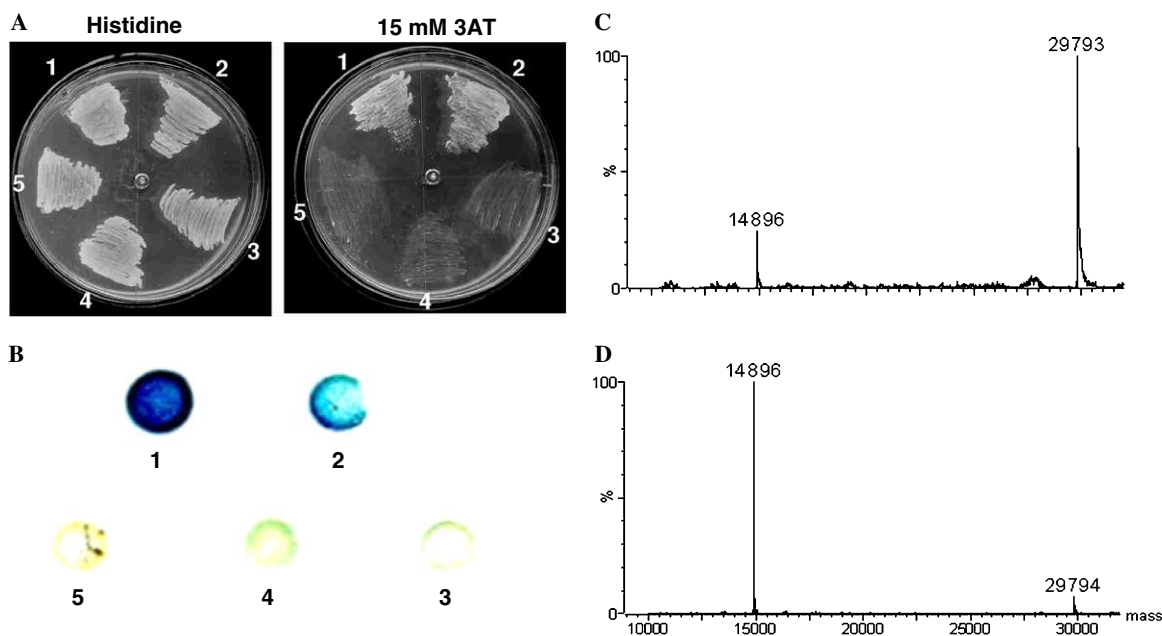


Fig. 6. SCAN domain dimerization analysis. (A,B) Yeast two-hybrid interaction assay on synthetic minimal medium (containing either histidine or 15 mM of 3-AT) and β -galactosidase filter assay, respectively. 1, Test strain containing the SCAN domain fused both to the DNA binding domain of *lexA* (DB-SCAN) and to the *GAL4* activation domain (AD-SCAN); 2, positive control (DB-Nip7/AD-Nop8); 3, 4, and 5 indicate the combinations DB-SCAN/AD, DB/AD-SCAN, and DB-Nop8/AD-SCAN used as negative controls. DB, DNA binding domain. AD, Activation domain. (C,D) Mass spectra of the SCAN domain in ammonium bicarbonate buffer and in ammonium bicarbonate buffer containing 50% acetonitrile, 0.1% formic acid, respectively.

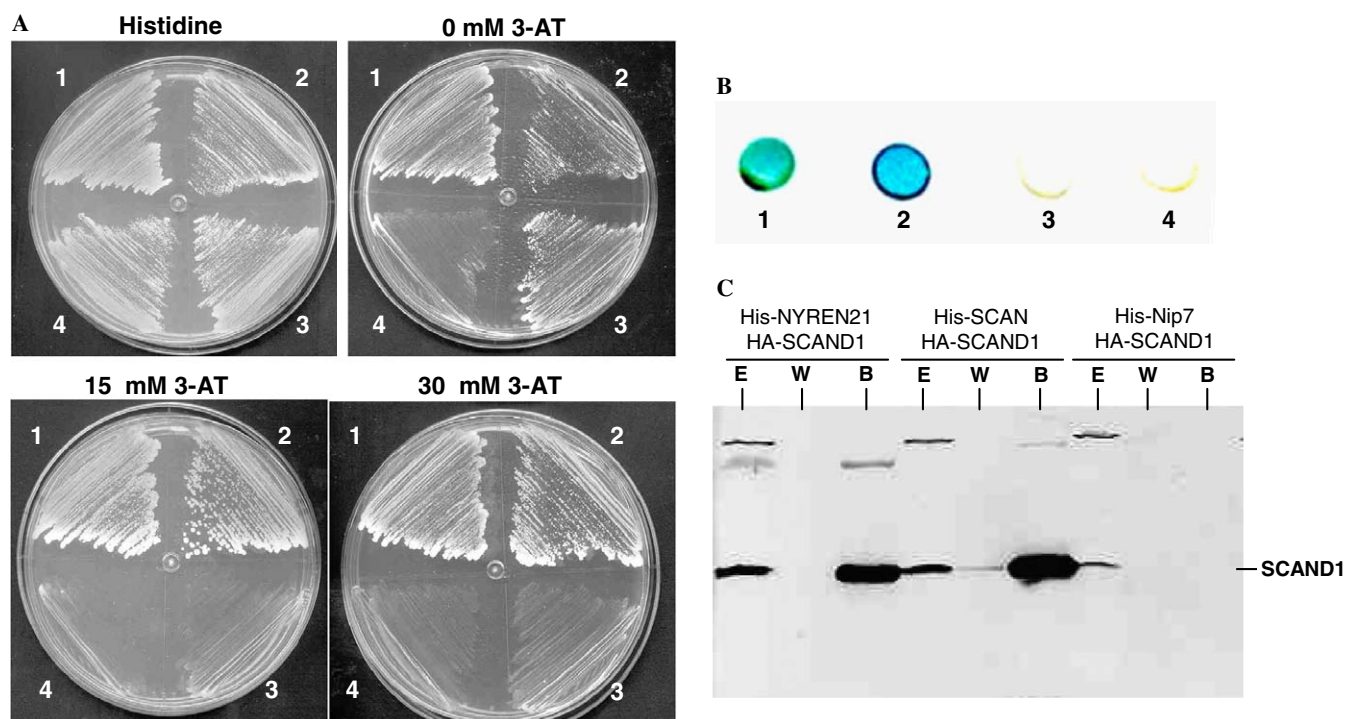


Fig. 7. Interaction between SCAN/NY-REN-21 and SCAND1. (A,B) Yeast two-hybrid interaction assay on synthetic minimal medium (containing histidine 0, 15 or 30 mM of 3-AT) and β -galactosidase filter assay, respectively. 1, Test strain containing the SCAN domain of NY-REN-21 fused to the DNA binding domain of *lexA* (DB-SCAN) and SCAND1 fused to the *GAL4* activation domain (AD-SCAND1); 2, positive control strain containing DB-SCAN/AD-SCAN; 3 and 4, negative control strains containing the combinations DB/AD-SCAND1 and DB-Nop8/AD-SCAN, respectively. (C) "Pull-down" assay of His-NY-REN-21/HA-SCAND1, His-SCAN/HA-SCAND1 and His-Nip7p/HA-SCAND1 (negative control). The histidine-tagged proteins were purified by affinity chromatography and the interacting HA-tagged SCAND1 (indicated) was detected by immunoblotting with an anti-HA antibody. Samples from the extract (E), the wash fraction (W) and from the resin-bound fraction eluted with imidazole (B) were fractionated by SDS-PAGE followed by immunoblot analysis.

Human SCAND1 and its mouse ortholog PGC-2 [30] are particular members of the SCAN domain protein family because they lack the zinc finger region. As a consequence, the heterodimer formed by NY-REN-21/SCAND1 is asymmetric concerning the DNA binding region since it contains just one zinc finger region and this asymmetric structure may have important implications on the control of gene expression by NY-REN-21. SCAND1 is expressed in a diverse number of tissues with higher levels detected in prostate, kidney, thyroid, liver, and thymus [9,31]. SCAND1 may play an important role in the mechanism of transcription regulation. In addition to NY-REN-21, SCAND1 interacts with the transcription factors MZF1B [7], ZNF202, ZNF191 [9], and with the nuclear receptors PPAR γ 2, α e δ e ER α [30]. Its interaction with PPAR γ 2 increased the trans-activation activity of PPAR γ 2 in transiently transfected cells [30]. Interestingly, SCAND1 and PGC-2 fused to the *GAL4* activation domain resulted in proteins unable to activate or repress reporter genes in transiently transfected HeLa cells. However, a truncated mutant containing the N-terminal region of SCAND1 and lacking the SCAN domain showed trans-activation activity suggesting that the SCAN domain hinders the trans-activation function of SCAND1 [30]. In conclusion, we have presented data showing that the SCAN domain of NY-REN-21 is able to form homodimers and asymmetric heterodimers with SCAND1. There are few reports in the literature on NY-REN-21 function and identifying the interactions among the SCAN family members is essential for the understanding of the regulatory mechanism used by this group of transcription factors to control gene expression.

Acknowledgments

This work was supported by Grant 00/02788-4 to N.I.T.Z. and the SMolBNet and CEPID/CBME programs from the Foundation for Research Support of the State of São Paulo (FAPESP). FRGC and THV are recipients of FAPESP and CNPq fellowships, respectively. We thank Zildene G. Correa for DNA sequencing, Thelma A. Pertinhez for helping with CD analysis, and Amadeu H. Iglesias for mass spectrometry analysis.

References

- [1] M.J. Scanlan, J.D. Gordan, B. Williamson, E. Stockert, N.H. Bander, V. Jongeneel, A.O. Gure, D. Jager, E. Jager, A. Knuth, Y.T. Chen, L.J. Old, Antigens recognized by autologous antibody in patients with renal-cell carcinoma, *Int. J. Cancer* 83 (1999) 456–464.
- [2] A.J. Williams, L.M. Khachigian, T. Shows, T. Collins, Isolation and characterization of a novel zinc-finger protein with transcription repressor activity, *J. Biol. Chem.* 270 (1995) 22143–22152.
- [3] J.H. Laity, B.M. Lee, P.E. Wright, Zinc finger proteins: new insights into structural and functional diversity, *Curr. Opin. Struct. Biol.* 11 (2001) 39–46.
- [4] V.J. Bardwell, R. Treisman, The POZ domain: a conserved protein–protein interaction motif, *Genes Dev.* 8 (1994) 1664–1677.
- [5] E.J. Bellefroid, D.A. Poncelet, P.J. Lecocq, O. Revelant, J.A. Martial, The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins, *Proc. Natl. Acad. Sci. USA* 88 (1991) 3608–3612.
- [6] T.L. Sander, K.F. Stringer, J.L. Maki, P. Szauter, J.R. Stone, T. Collins, The SCAN domain defines a large family of zinc finger transcription factors, *Gene* 310 (2003) 29–38.
- [7] T.L. Sander, A.L. Haas, M.J. Peterson, J.F. Morris, Identification of a novel SCAN box-related protein that interacts with MZF1B. The leucine-rich SCAN box mediates hetero and homoprotein associations, *J. Biol. Chem.* 275 (2000) 12857–12867.
- [8] A.J. Williams, S.C. Blacklow, T. Collins, The zinc finger-associated SCAN box is a conserved oligomerization domain, *Mol. Cell. Biol.* 19 (1999) 8526–8535.
- [9] C. Schumacher, H. Wang, C. Honer, W. Ding, J. Koehn, Q. Lawrence, C.M. Coulis, L.L. Wang, D. Ballinger, B.R. Bowen, S. Wagner, The SCAN domain mediates selective oligomerization, *J. Biol. Chem.* 275 (2000) 17173–17179.
- [10] J.R. Stone, J.L. Maki, S.C. Blacklow, T. Collins, The SCAN domain of ZNF174 is a dimer, *J. Biol. Chem.* 277 (2002) 5448–5452.
- [11] C. Honer, P. Chen, M.J. Toth, C. Schumacher, Identification of SCAN dimerization domains in four gene families, *Biochim. Biophys. Acta* 1517 (2001) 441–448.
- [12] D. Ivanov, J.R. Stone, J.L. Maki, T. Collins, G. Wagner, Mammalian SCAN domain dimer is a domain-swapped homolog of the HIV capsid C-terminal domain, *Mol. Cell* 17 (2005) 137–143.
- [13] K. Chowdhury, M. Goulding, C. Walther, K. Imai, H. Fickenscher, The ubiquitous transactivator Zfp-38 is upregulated during spermatogenesis with differential transcription, *Mech. Dev.* 39 (1992) 129–142.
- [14] X.W. Yang, R. Zhong, N. Heintz, Granule cell specification in the developing mouse brain as defined by expression of the zinc finger transcription factor RU49, *Development* 122 (1996) 555–566.
- [15] X.W. Yang, C. Wynder, M.L. Doughty, N. Heintz, BAC-mediated gene-dosage analysis reveals a role for Zipro1 (Ru49/Zfp38) in progenitor cell proliferation in cerebellum and skin, *Nat. Genet.* 22 (1999) 327–335.
- [16] P. James, J. Halladay, E.A. Craig, Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast, *Genetics* 144 (1996) 1425–1436.
- [17] P.L. Bartel, C. Chien, R. Sternglanz, S. Fields, Using the two-hybrid system to detect protein–protein interactions, in: D.A. Hartleu (Ed.), *Cellular Interactions in Development: A Practical Approach*, IRL Press, Oxford, 1993, pp. 153–179.
- [18] S.M. Hollenberg, R. Sternglanz, P.F. Cheng, H. Weintraub, Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system, *Mol. Cell. Biol.* 15 (1995) 3813–3822.
- [19] N.I.T. Zanchin, D.S. Goldfarb, Nip7p interacts with Nop8p, an essential nucleolar protein required for 60s ribosome biogenesis, and the exosome subunit rrp43p, *Mol. Cell. Biol.* 19 (1999) 1518–1525.
- [20] M.S. Lee, G.P. Gippert, K.V. Soman, D.A. Case, P.E. Wright, Three-dimensional solution structure of a single zinc finger DNA-binding domain, *Science* 245 (1989) 635–637.
- [21] M.P. Foster, D.S. Wuttke, I. Radhakrishnan, D.A. Case, J.M. Gottesfeld, P.E. Wright, Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA, *Nat. Struct. Biol.* 4 (1997) 605–608.
- [22] T. Matt, M.A. Martinez-Yamout, H.J. Dyson, P.E. Wright, The CBP/p300 TAZ1 domain in its native state is not a binding partner of MDM2, *Biochem. J.* 381 (2004) 685–691.
- [23] D.T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, *J. Mol. Biol.* 292 (1999) 195–202.
- [24] P. Romero, Z. Obradovic, X. Li, E.C. Garner, C.J. Brown, A.K. Dunker, Sequence complexity of disordered protein, *Proteins* 42 (2001) 38–48.

- [25] V.S. Alves, B.A. Castilho, Gir2 is an intrinsically unstructured protein that is present in *Saccharomyces cerevisiae* as a group of heterogeneously electrophoretic migrating forms, *Biochem. Biophys. Res. Commun.* 332 (2005) 450–455.
- [26] K. Bilecen, U.H. Ozturk, A.D. Duru, T. Sutlu, M.V. Petoukhov, D.I. Svergun, M.H. Koch, U.O. Sezerman, I. Cakmak, Z. Sayers, Triticum durum metallothionein. Isolation of the gene and structural characterization of the protein using solution scattering and molecular modeling, *J. Biol. Chem.* 280 (2005) 13701–13711.
- [27] H.J. Dyson, P.E. Wright, Unfolded proteins and protein folding studied by NMR, *Chem. Rev.* 104 (2004) 3607–3622.
- [28] A.L. Fink, Natively unfolded proteins, *Curr. Opin. Struct. Biol.* 15 (2005) 35–41.
- [29] D.M. Jacobs, A.S. Lipton, N.G. Isern, G.W. Daughdrill, D.F. Lowry, X. Gomes, M.S. Wold, Human replication protein A: global fold of the N-terminal RPA-70 domain reveals a basic cleft and flexible C-terminal linker, *J. Biomol. NMR* 14 (1999) 321–331.
- [30] R. Babb, B.R. Bowen, SDP1 is a peroxisome-proliferator-activated receptor gamma 2 co-activator that binds through its SCAN domain, *Biochem. J.* 370 (2003) 719–727.
- [31] T.L. Sander, J.F. Morris, Characterization of the SCAN box encoding RAZ1 gene: analysis of cDNA transcripts, expression, and cellular localization, *Gene* 296 (2002) 53–64.

4- DISCUSSÃO

4.1- Triagem para a identificação dos ligantes protéicos da proteína INT6 humana

Desde a identificação do gene que codifica a proteína INT6 como o sítio de inserção do genoma do MMTV, esta proteína tem sido alvo de estudos por diversos pesquisadores. A integração do genoma viral pode levar à expressão de proteínas truncadas sem a porção C-terminal (Marchetti *et al.*, 1995), levantando a questão de que as formas truncadas poderiam atuar como mutantes dominantes negativos, ganhar função ou acarretar a perda da função da INT6. Miyazaki *et al.* (1997) e Buttitta *et al.* (2005), estudando carcinomas primários de mama e de pulmão, respectivamente, sustentam que a perda de função da INT6 pode estar relacionada com a transformação maligna das células. Por outro lado, dois outros trabalhos (Rasmussen *et al.*, 2001; Mayeur & Hershey, 2002) mostraram que a forma truncada hINT6T1139, que compreende os aminoácidos 1 a 157, levou à transformação das células mesmo na presença de dois alelos normais do gene, sugerindo que esta forma truncada atue como um mutante dominante negativo.

A proteína INT6 apresenta um sinal de localização nuclear e um sinal de exportação nuclear tendo sido identificada tanto no núcleo quanto no citoplasma (Diella *et al.*, 1997; Guo & Sen, 2000; Watkins & Norbury, 2004). Além disso, a proteína INT6 é a subunidade de 48 kDa do eIF3 e co-purifica com os complexos protéicos CSN e proteassomo. Estas características indicam que a proteína INT6 pode ter diferentes funções e ligantes protéicos. Visando esclarecer a função da proteína INT6, foram realizadas 3 triagens pelo método do duplo-híbrido de levedura usando como iscas a proteína hINT6, a porção C-terminal que inclui o domínio PCI e a forma truncada hINT6T22.

O método do duplo-híbrido de levedura consiste na expressão de 2 proteínas de fusão, uma contendo um domínio de ligação ao DNA e outra contendo um domínio de ativação da transcrição. Estas proteínas são co-expressas em uma cepa de levedura com genes repórteres cujos promotores apresentam seqüências para a

ligação específica do domínio de ligação ao DNA usado (Fields & Song, 1989). Quando as proteínas de fusão interagem ocorre a ativação dos genes repórteres. Apesar deste método ser amplamente usado e apresentar como vantagens ser um método com baixo custo financeiro, gerar resultados rapidamente e testar as interações *in vivo*, as principais desvantagens são o isolamento de interações falso-positivas e a ativação dos genes repórteres do sistema pela proteína alvo do estudo (Vidal & Legrain, 1999; Fields, 2005).

As interações falso-positivas são aquelas que quando fusionadas ao domínio de ligação ao DNA ou ao domínio de ativação da transcrição ativam os genes repórteres independentemente de interação. Frequentemente estas proteínas são fatores de transcrição que contêm domínios que ativam os genes repórteres quando em fusão ao domínio de ligação ao DNA. Outras proteínas se comportam como falso-positivos por razões ainda não esclarecidas (Vidal & Legrain, 1999). É estimado que cerca de 50% ou mais das interações encontradas apenas com o método do duplo-híbrido sejam falso-positivas (von Mering *et al.*, 2002). Portanto, após a identificação das seqüências é necessária uma análise criteriosa dos resultados e a confirmação das interações por outra metodologia.

Tanto a proteína hINT6, como o domínio PCI usados como iscas ativaram os genes repórteres do sistema. Desta forma foi adicionado 3-aminotriazol (3-AT) nas placas de seleção nas concentrações de 30 e 10 mM, respectivamente. O 3-AT atua como um inibidor competitivo da enzima codificada pelo gene *HIS3*, impedindo a síntese de histidina e, conseqüentemente, o crescimento da levedura. Sob estas condições, é esperado que a interação entre as proteínas de fusão leve a um aumento da expressão de *HIS3* que seja capaz superar o efeito causado pelo 3-AT e, conseqüentemente, ao crescimento e seleção das colônias.

As triagens pelo método do duplo-híbrido usando a proteína hINT6 e o domínio PCI como iscas isolaram muitas proteínas, num total de 52 proteínas diferentes com funções distintas (Tabela 2, item 3.1) sendo 15 ainda não caracterizadas. As proteínas que estavam *in frame* com o domínio de ativação da transcrição *GAL4* foram analisadas através do programa SMART (Schultz *et al.*, 1998, 2000) para a identificação de domínios protéicos conhecidos. Proteínas

contendo domínios de ligação a DNA, RNA e domínios de interação entre proteínas representaram a maioria dos clones, somando 20 seqüências. As demais seqüências apresentaram outros domínios ou não apresentaram domínios conhecidos até o momento.

Entre as proteínas isoladas usando a proteína hINT6 como isca, estão antígenos encontrados em determinados tumores como as proteínas NY-REN-21 (clones I5.67 e I5.68) e NY-REN-41 (clone I5.100), encontradas em indivíduos com carcinoma de células renais (Scanlan *et al.*, 1999) e o antígeno se70-2 (clone I4.5) de linfoma cutâneo de células T (Eichmüller S, 2001).

A proteína Rab11b (clones I4.48, I5.15, I5.30), membro da família de proteínas do oncogene RAS, além das proteínas fosfoproteína de fase M-8 (MPP8, clones I4.13 e I5.1, Matsumoto-Taniura *et al.*, 1996) de função ainda não caracterizada e Rent-1, envolvida na degradação de mRNA com mutações, (clone I4.8, Sun *et al.*, 1998) também foram identificadas nesta triagem.

Proteínas que interagem com elementos do citoesqueleto foram identificadas usando o domínio PCI como isca (clones P4.43 e P5.2). O clone P4.43 codifica uma proteína ainda não caracterizada que apresenta o domínio homólogo a calponina (CH, *Calponin homolog domain*) de interação com actina. Este domínio é encontrado tanto em proteínas do citoesqueleto, quanto em proteínas envolvidas em transdução de sinais (Gimona *et al.*, 2002). O clone P5.2 codifica a cadeia α de tubulina. É interessante considerar que cepas de *S. pombe* Δ int6 apresentam defeitos na segregação dos cromossomos por modificação na polimerização e despolimerização dos microtúbulos (Yen & Chang, 2000). Além disso, a proteína INT6 de *S. pombe* forma um complexo com a proteína Moe 1 atuando na segregação correta dos cromossomos (Yen & Chang, 2000). Acredita-se que a proteína INT6 atue na polimerização e despolimerização dos microtúbulos indiretamente por interação com outras proteínas. Até o momento não foram publicados experimentos demonstrando interação direta da INT6 com microtúbulos.

Os clones I6.91 e P6.10 são de especial interesse pois codificam as proteínas eIF3a/p170 e Ret finger, cujas interações com a hINT6 já foram descritas por Asano *et al.* (1997) e Morris-Desbois *et al.* (1999), respectivamente. A proteína eIF3a/p170

faz parte do eIF3 e sua superexpressão foi identificada em cânceres de mama (Bachmann *et al.*, 1997), cervical (Dellas *et al.*, 1998), esôfago (Chen & Burger, 1999) e pulmão (Pincheira *et al.*, 2001), sugerindo que esta proteína possa estar envolvida em controle do ciclo celular. A proteína Ret finger é um repressor transcricional e pertence à família de proteínas *B-box RING finger* (Shimono *et al.*, 2000). As proteínas desta família participam de diversos processos dentro da célula, como oncogênese, apoptose e ubiquitinação (Topcu *et al.*, 1999; Borden, 2000). Outras duas proteínas contendo *RING finger* foram identificadas. Uma delas é uma proteína não caracterizada (clone P5.28) e a outra é a proteína Rbx 1 (clone I5.34), parte do complexo ubiquitina ligase (Kamura *et al.*, 1999).

O isolamento das proteínas eIF3a/170 e Ret finger revelaram que, apesar do grande número de seqüências diferentes e das dificuldades técnicas para a exclusão das proteínas que poderiam atuar como falso positivos, as triagens usando a proteína INT6 e o domínio PCI como iscas foram capazes de isolar interações específicas. Desta forma, considerando os resultados dos testes dos genes repórteres, assim como tamanho do cDNA no vetor pACT2 foram selecionadas 4 proteínas para a validação da interação com a hINT6 por ensaios *in vitro*. O fato destas proteínas poderem estar envolvidas em transdução do sinal e controle do ciclo celular também foi bastante relevante para a escolha das mesmas.

As proteínas selecionadas, NY-REN-21, I κ BL, alfa4 e 14-3-3 τ , foram clonadas em fusão com cauda de histidina, apresentando boa expressão na fração solúvel do extrato de *E. coli*, com exceção da proteína I κ BL. A proteína hINT6 foi clonada em fusão à cauda de histina, proteína A e Glutathione S-transferase (GST). Em todos os casos a expressão da hINT6 na fração solúvel foi muito baixa dificultando bastante os testes de interação *in vitro*. Como forma de concluir os testes de interação, a tradução *in vitro* da proteína hINT6 também foi realizada. Diversas tentativas para a confirmação das interações foram testadas, entretanto sem sucesso. Apesar dos resultados não terem sido conclusivos, 2 proteínas isoladas nestas triagens, as proteínas alfa4 e Rab11b, foram estudadas em mais detalhes pelas alunas Juliana H. C. Smetana e Sandra M. N. Scapin, respectivamente, e resultaram em duas publicações (Anexos 1 e 2).

Acreditamos que as dificuldades encontradas nesta etapa do projeto estejam relacionadas à ativação dos genes repórteres do sistema de duplo-híbrido pela proteína hINT6 e domínio PCI, revelando que esta metodologia não foi adequada para a identificação dos ligantes protéicos destas proteínas. O fato da proteína hINT6 não expressar de forma solúvel no extrato de *E. coli* também dificultou os testes de interação *in vitro*.

Com base nestes resultados buscamos a identificação de uma porção da proteína hINT6 que não fosse capaz de ativar os genes repórteres. Para isto clonamos no vetor pTL-1 as formas truncadas hINT6T1139 e hINT6T22. Estas proteínas não ativaram os genes *HIS3* e *lacZ* e suas expressões, em fusão com o domínio *lexA* foram identificadas com sucesso nos extratos da levedura L40 por *immunoblot* (Figura 2, item 3.1). Escolhemos a forma truncada hINT6T22 para a realização do terceiro duplo-híbrido, que se mostrava bastante promissor visto que poderiam ser encontradas interações que indicassem uma explicação para a transformação maligna das células.

Neste duplo-híbrido foram isoladas principalmente as proteínas PLZF, HIRIP3 e Histona H3. Identificamos que todas estas proteínas ativaram os genes repórteres na ausência da isca hINT6T22 (Figura 3, item 3.1), mostrando que a ativação dos genes repórteres foi inespecífica, não sendo resultado de interação entre as proteínas de fusão. É possível que a proteína de fusão *lexA*-hINT6T22 não tenha enovelado corretamente ou não tenha translocado para o núcleo. Estes dois eventos poderiam explicar porque não identificamos nenhuma interação específica. Uma terceira hipótese seria que a forma truncada hINT6T22 não seja funcional e perdeu a capacidade de interação com outras proteínas. Isto pode ser sustentado pela ausência do domínio PCI nesta proteína, que tem sido relatado como um domínio de interação entre proteínas (Yen & Chang, 2000; Kim *et al.*, 2001; Yahalom *et al.*, 2001).

Yen & Chang (2000) observaram que a proteína INT6 interage com a proteína Moe 1 e que as formas truncadas hINT6T22 e hINT6T1139 perdem a capacidade de interação. Além disso, observaram que a proteína humana INT6, mas não suas formas truncadas, é capaz de recuperar os defeitos de crescimento de células de *S.*

pombe Δ int6, sugerindo que as duas formas truncadas encontradas em tumores mamários de camundongos perdem a função da INT6. Os dados estruturais descritos neste trabalho mostram que o domínio PCI pode formar um domínio estrutural resistente à proteólise e são coerentes com os dados do duplo-híbrido que indicam que a INT6 sem a porção que contém o domínio PCI não é funcional.

4.2- Análise da proteína INT6 de *A. thaliana* por dicroísmo circular (CD)

A proteína hINT6 recombinante se mostrou totalmente insolúvel quando expressa em extratos de *E. coli*. Induções da proteína hINT6 fusionada à GST, proteína A e cauda de histidina foram realizadas em diferentes temperaturas, entretanto houve pouca diferença em relação à solubilidade destas proteínas (dados não mostrados), o que impossibilitou estudos estruturais com a proteína hINT6. A proteína INT6 é bastante conservada, apresentando ortólogos com alta similaridade de seqüências de aminoácidos em diferentes espécies (Marchetti *et al.*, 1995; Desbois *et al.*, 1996; Miyazaki *et al.*, 1999; Morris-Desbois *et al.*, 1999; Bandyopadhyay *et al.*, 2000; Crane *et al.*, 2000; Yen & Chang, 2000; Akiyoshi *et al.*, 2001; Yahalom *et al.*, 2001). A proteína INT6 de *A. thaliana* apresenta 62% de similaridade com a proteína humana (Figura 4.1).

Testes de expressão da proteína AtINT6 fusionada à cauda de histidina (His-AtINT6) revelaram que esta proteína é parcialmente solúvel em extratos de *E. coli* quando induzida a 25°C. A proteína His-AtINT6 foi purificada por cromatografia de afinidade, seguida por cromatografia de troca iônica, obtendo-se a proteína em bom grau de pureza para os ensaios estruturais (Figura 4C, artigo I). Visando determinar se a proteína INT6 apresenta regiões mais compactas e principalmente, se o domínio PCI definido por alinhamento da estrutura primária, corresponde a um domínio estrutural, foram realizados experimentos de proteólise limitada com proteases de diferentes especificidades, seguidos de análises por espectrometria de massas para a identificação das regiões resistentes. A técnica de proteólise limitada é muito usada para a identificação de regiões pouco enoveladas nas proteínas, uma vez que

domínios compactos apresentam maior resistência à digestão por proteases do que domínios flexíveis (Fontana *et al.*, 1997; Iakoucheva *et al.*, 2001).

O resultado das análises por proteólise limitada e espectrometria de massas revelaram que os fragmentos resistentes incluem o domínio PCI. Os produtos de proteólise das enzimas tripsina e termolisina apresentam massa molecular semelhante, mas se diferenciam nas regiões N e C-terminais (Figura 1, artigo I). Com base nestes resultados, construímos 3 proteínas truncadas diferentes AtINT6TR1 (aminoácidos 124 a 441), AtINT6TR2 (aminoácidos 124 a 415) e AtINT6TR3 (aminoácidos 172 a 415).

A proteína AtINT6TR1 foi totalmente insolúvel quando expressa em *E. coli*, sugerindo que os 26 aminoácidos da região C-terminal da proteína interferem negativamente em seu enovelamento (Figura 4A e B, artigo I). A proteína AtINT6TR2 apresenta bom nível de expressão na fração solúvel do extrato de *E. coli* quando induzida a 25°C (Figura 4A, artigo I). Entretanto, o melhor nível de expressão e solubilidade foi obtido com a indução da proteína AtINT6TR3 (Figura 4A e B, artigo I). É importante ressaltar que esta forma truncada apresenta as extremidades N e C terminais mais próximas dos limites do domínio PCI da proteína AtINT6 (aminoácidos 288 a 404) preditos por análises computacionais. A purificação das proteínas AtINT6TR2 e AtINT6TR3 foi obtida com sucesso e com bom grau de pureza para os ensaios de dicroísmo circular.

A técnica de espectropolarimetria de dicroísmo circular se baseia na diferença de absorção da luz circularmente polarizada à esquerda e à direita pelas moléculas quirais. Moléculas quirais são aquelas que não são superponíveis com sua imagem no espelho e representam quase todas as moléculas biológicas, incluindo as proteínas. A técnica de CD é usada para estimar a quantidade de estrutura secundária de proteínas, assim como estudar efeitos de mudanças conformacionais causadas por ligantes, agentes desnaturantes e temperatura (Woody, 1995; Kelly & Price, 1997).

Até o momento não existem relatos na literatura de estudos estruturais com a proteína INT6. As análises de dicroísmo circular das proteínas AtINT6, AtINT6TR2 e AtINT6TR3 revelaram que todas apresentam alto conteúdo de α -hélice (Figura 5,

artigo I). Este resultado está de acordo com a predição de estrutura secundária da proteína AtINT6 (Figura 2, artigo I), realizada pelo programa PSIPRED (Jones *et al.*, 1999; McGuffin *et al.*, 2000) e com as predições realizadas para o domínio PCI (Hofmann & Bucher, 1998; Kim *et al.*, 2001). A forma AtINT6TR3 apresentou o espectro de dicroísmo circular com maior intensidade de sinal, sugerindo maior quantidade de α -hélice que a proteína AtINT6 e AtINT6TR2, cujos espectros foram bastante parecidos. Além disso, houve um deslocamento do mínimo de 222 nm, característico de proteínas ricas em estrutura do tipo α -hélice, para 225 nm. Subtraindo-se o espectro da proteína AtINT6TR3 pelo espectro da proteína AtINT6, verifica-se um mínimo em 230 nm, que pode estar relacionado ao menor conteúdo de resíduos aromáticos da proteína AtINT6TR3, comparado com as proteínas AtINT6 e AtINT6TR2, sendo de 28, 46 e 37 aminoácidos, respectivamente (Figura 5, artigo I). Os resíduos aromáticos, fenilalanina, tirosina e triptofano, apresentam contribuição de sinal positivo no espectro de CD na região de 230 nm (Sreerama *et al.*, 1999).

Visto que a proteína AtINT6TR3 apresentou maior nível de expressão e solubilidade, análises de desnaturação térmica, acompanhando o sinal a 222 nm, foram realizadas para verificar se esta forma truncada seria mais estável que as proteínas AtINT6 e AtINT6TR2. Os dados obtidos foram normalizados para uma escala de 0 a 1, de forma que o sinal obtido com a proteína mais enovelada (menor temperatura) corresponde a 0 e o sinal obtido com a proteína menos enovelada corresponde a 1. A temperatura média de transição (T_m) do estado mais enovelado para o menos enovelado da proteína foi de 44°C, 49°C e 58°C para as proteínas AtINT6, AtINT6TR2 e AtINT6TR3, respectivamente. A T_m da proteína AtINT6TR3 foi a mais elevada, confirmando que esta forma truncada é mais estável que as demais. Acreditamos que a diferença de solubilidade e estabilidade entre as proteínas AtINT6TR2 e AtINT6TR3 pode ser devido à presença de maior número de resíduos hidrofóbicos entre os aminoácidos 124 a 172, além de mais 2 cisteínas. Não entendemos como os 26 aminoácidos da região C-terminal pode ter tanta influência na solubilidade da proteína AtINT6TR1.

A identificação de uma forma truncada solúvel e estável que compreenda o domínio PCI pode ser muito útil para a determinação da estrutura tridimensional deste

domínio. A caracterização estrutural do domínio PCI é importante para ajudar a entender sua função, visto que as formas truncadas da proteína INT6 perdem este domínio. Além disso, a região de interação da proteína INT6 com outras proteínas parece ser o domínio PCI (Yen & Chang, 2000; Yahalom *et al.*, 2001).

As proteínas que apresentam o domínio PCI parecem ter participação na montagem correta dos complexos protéicos que fazem parte (Freilich *et al.*, 1999; Tsuge *et al.*, 2001). Em eucariotos existem pelo menos 3 complexos multiprotéicos distintos que têm sido chamados de complexos PCI (Kim *et al.*, 2001): o proteassomo, o complexo COP9/signalossomo e o fator 3 de iniciação da tradução. Como característica comum, estes complexos são formados por várias subunidades que apresentam o domínio PCI (Hofmann & Bucher, 1998).

Um estudo recente de alinhamento de seqüências usando ferramentas de bioinformática revelou indícios de que o domínio PCI consiste de dois domínios estruturalmente distintos: a porção C-terminal mais conservada e com limites bem definidos e a N-terminal menos conservada e com limites pouco definidos (Scheel & Hofmann, 2005). Neste estudo foi observado que as proteínas perdem a conservação da seqüência do domínio PCI em diferentes posições da porção N-terminal. Um caso extremo é a proteína INT6 onde somente a porção C-terminal do domínio foi encontrada. Este resultado pode ser uma indicação para a independência funcional e estrutural das sub-regiões N e C-terminais do domínio PCI e reforça a importância da elucidação da estrutura do domínio PCI da proteína INT6, visto que representa, até o momento, uma forma única do domínio.

```

A.thaliana      MEEKSQNYDLTPLIAPNLDRLVFP IFEFLQERQLYPDEQILKSKIQLLNQTNMVDYAMD
humana          ----MAEYDIT TR IAHFLDRHLVFP LLEFLSVKE IYNEKELLQGKLDLL SD TNMVD FAMD
                :****. *  *****:***. :*: :*: :*: :*: :*: :*: :*
A.thaliana      IHKSLYHTEDAPQEMVERRTEVVARLK SLE EAAA PLVSF LLNPN AVQELR--ADKQYNLQ
humana          VYKNLYS-DDI PHALREKRTTVVAQLQLQAETEP IVKMFEDPETRQMQSTRDGRMLFD
                :*. *  : * : : * : *  * : * : : * : : : : * : :
A.thaliana      MLKERYQIGPDQI EALYQ YAKF QFE CGNYSGAAD YLY QY RT LCS NLER-SLSALWGKLAS
humana          YLADKHGFRQEYLDTLRYAKFQYECGNYSGAAEYLYFFRVLPATDRNALSSLWGKLAS
                * : : : : : : : * : * : * : * : * : * : * : * : * : * : * :
A.thaliana      E ILMQNWD IAL EE LNRLKEI DSKS FS SPLNQ VQNRI WLMHWGLYI FFDH NGRTQ II DL
humana          E ILMQNWD AAMEDLTRLKEITIDNNSVSSPLQSLQORTWLIHWSLFVFFNHPKGRDNIIDL
                ***** *: * : . ***** * : . * : * : : : * : * : * : * : * : * :
A.thaliana      FNQD-KYLNAIQTSA PHLLRYLATAFIVNK ---RRRPQLKEFIKVIQQEHYSYKDP IIEF
humana          FLYQPQYLNAIQTMCPHILRYLT TAVI TNKDVRKRRQVLKDLVKVIQQESYTYKDP IIEF
                * : : ***** . * : * : * : * : * : * : * : * : * : * : * : * :
A.thaliana      LACVFNVD FDGA QKKMKECEEVIVND PFLGKRV EDGNF STVPLRD EFL ENARLFVFE TY
humana          VECLYVNFDFDGAQKKLRECESVLVNDFFLVACLED -----F IENARLFIETEF
                : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :
A.thaliana      CKIHQRIDMGV LA EKLNL NYEEAERWI VNL IR TS KID AK IDSES GTVIMEP TQPNV HE QL
humana          CRIHQCI SINMLADKLNMTPEEAERWIVNLIRNARLDAKIDSKLGHVVMGNNAVSPYQQV
                * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :
A.thaliana      INHTKGLSGRTYKLVNQLLEHTQAQATR-----
humana          IEKTKSLSFRSQMLAMNIEKKLNQNSRSEAPNWTQDSGFY
                * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :

```

Figura 4.1: Alinhamento da sequência de aminoácidos das proteínas hINT6 e AtINT6.

4.3- Análises da proteína NY-REN-21

O antígeno NY-REN-21 foi identificado por Scanlan *et al.* (1999) em pacientes com carcinoma renal pelo método de SEREX. Desde então nenhum estudo foi publicado a respeito da estrutura e função desta proteína, que representa uma possível ortóloga para a proteína ZFP38 de camundongo com 82% de identidade. Em camundongos a ZFP38 se liga à sequência consenso de DNA 5'AGTAC3' (Yang *et al.*, 1996) e contém 3 seqüências *in tandem* no N-terminal que ativam a transcrição de genes repórteres quando fusionadas ao domínio *GAL4* de ligação ao DNA.

Tanto a proteína humana NY-REN-21 quanto a proteína ZFP38 apresentam sete domínios dedos de zinco em sua porção C-terminal e um domínio SCAN, na

porção N-terminal. Estas duas regiões são ligadas por uma região de cerca de 150 aminoácidos predita como desordenada.

4.3.1- Proteína NY-REN-21: purificação e dicroísmo circular

A proteína NY-REN-21 foi clonada em fusão à cauda de histidina e sua expressão induzida por 6 horas na cepa de *E. coli* BL21(DE3) Δ *slyD*, resultando na expressão de boa quantidade de proteína na forma solúvel. O extrato de *E. coli* foi tratado com DNaseI antes do início da purificação, uma vez que análises anteriores mostraram que grande quantidade de DNA co-eluiu com a proteína NY-REN-21. Obtivemos a proteína em bom grau de pureza após a realização da purificação em duas etapas. Inicialmente o extrato foi aplicado em coluna de afinidade a níquel (HiTrap Chelating HP) para permitir a interação com a cauda de histidina. As frações mais puras foram injetadas em uma coluna de Heparina e foi aplicado um gradiente de 300 a 1,5 M de NaCl. A proteína eluiu pura com 900 mM do sal. Após a purificação, medidas de dicroísmo circular foram realizadas com o objetivo de caracterizar a quantidade de estrutura secundária da proteína NY-REN-21 expressa em *E. coli* e verificar o efeito do tratamento da proteína com EDTA na estrutura dos dedos de zinco. As análises foram realizadas no tampão de eluição da coluna de heparina. A quantidade elevada de sal se justifica pelo fato da proteína precipitar quando realizamos diálises para sua remoção ou diminuição da concentração. A amostra foi dividida em duas alíquotas e em uma delas foi adicionado 2 mM de EDTA 30 minutos antes da medida.

Foi observado que a proteína apresenta grande quantidade de estrutura desordenada em sua estrutura, provavelmente resultado da contribuição da região central (discutido abaixo) e da região dos 7 dedos de zinco (abrange aproximadamente os aminoácidos 275 a 473) e que, provavelmente, está parcialmente enovelada devido à ausência do DNA ligante específico (Figura 2, artigo II). Os dedos de zinco apresentam grande mobilidade nestas circunstâncias porque as

seqüências conservadas de aminoácidos que conectam os motivos C₂H₂ estão desestruturadas (Laity *et al.*, 2000b).

O tratamento com EDTA resultou em perda de estrutura secundária da proteína, provavelmente devido ao desenovelamento dos dedos de zinco. Foi observada a diminuição da leitura a 222 nm e deslocamento do mínimo de cerca de 205 nm em direção a 200 nm. Subtraindo-se os espectros (espectro da proteína nativa subtraído do espectro da proteína tratada com EDTA) observa-se um espectro típico de estrutura em α -hélice com mínimos em 208 e 222 nm e máximo em 190 nm (Figura 2, artigo II). Este resultado está de acordo com Matt *et al.* (2004) que estudando o efeito da adição de EDTA no domínio ligador de zinco TAZ1 da proteína CPB/p300 verificaram um espectro de CD com redução do conteúdo de α -hélice, quando comparado ao espectro da proteína nativa. Este resultado também mostra que a proteína recombinante expressa em *E. coli* apesar de estar em solução contendo quantidade elevada de NaCl mantém sua estrutura, que é alterada apenas com a adição do quelante de zinco.

4.3.2- Análise da região central desordenada da proteína NY-REN-21 (DCR-Disordered Central Region)

A região central da proteína NY-REN-21 (aminoácidos 128 a 276) é rica em aminoácidos polares e carregados (Figura 1, artigo II). Esta característica é comum a proteínas intrinsecamente desenoveladas que apresentam baixa hidrofobicidade e carga líquida alta, sendo os aminoácidos glutamina, lisina, arginina, glicina, serina, prolina e alanina predominantes (Dunker *et al.*, 2001; Romero *et al.*, 2001; Fink, 2005). Esta característica é usada como um dos parâmetros de análise nos programas computacionais de predição de regiões intrinsecamente desordenadas (PONDRs, *Predictors of Natural Disordered Regions*, Dunker *et al.*, 2001). De fato, a análise da proteína NY-REN-21 no programa PONDR (Romero *et al.*, 2001) e PSIPRED (Jones *et al.*, 1999; McGuffin *et al.*, 2000) revelaram que a sua região

central é predita como desordenada (Figura 4.2) e rica em estrutura tipo *random coil* (dados não mostrados), respectivamente.

A ocorrência de regiões intrinsecamente desenoveladas é bastante comum em proteínas eucarióticas (Dunker *et al.*, 2002; Uversky, 2002). Um estudo de predição de regiões desordenadas realizado por Ward *et al.*, (2004) revelou que 33% das proteínas eucarióticas apresentam longas regiões de desordem, contra 4,2% em eubactéria e 2,0% em arqueobactéria.

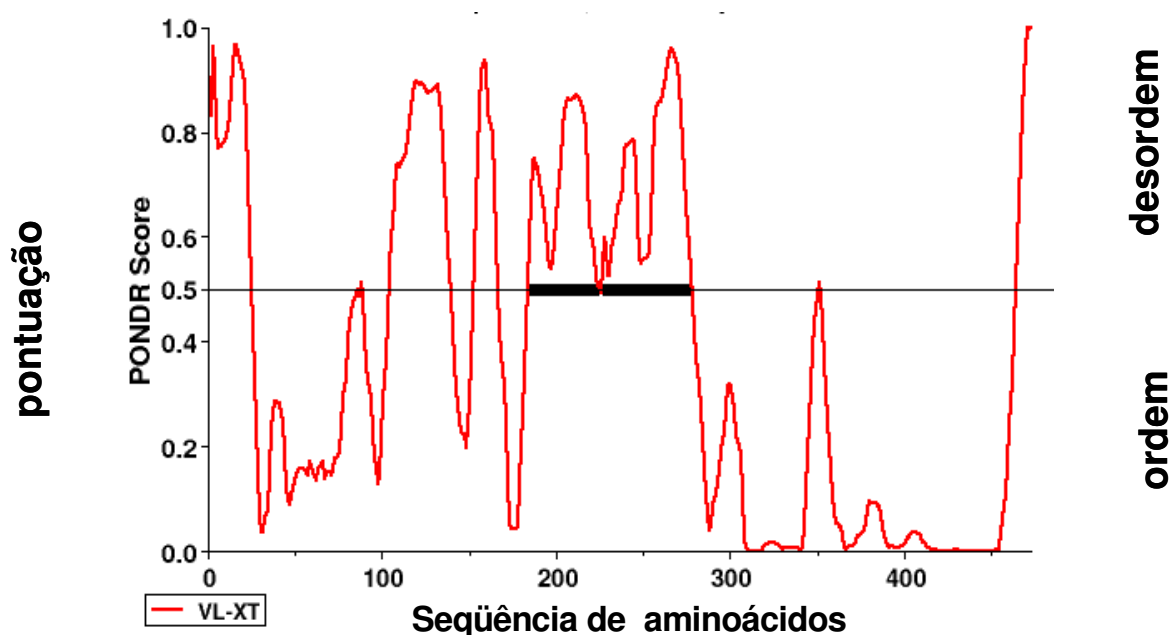


Figura 4.2: Predição de regiões desordenadas na proteína NY-REN-21 pelo programa PONDR. O grau de desordem analisado pelo programa é diretamente proporcional ao aumento da pontuação de 0 a 1, sendo considerado estrutura não ordenada os resíduos de aminoácidos com pontuação acima de 0,5. VL-XT: V- *variously characterized*; L- *long regions of disorder*; X- *X-ray characterized data*; T- *location at the termini*.

De modo a caracterizar e investigar a estrutura secundária desta região, os aminoácidos 128 a 276 foram clonados no vetor de expressão de *E. coli* pETGST-TEV. A proteína de fusão GST-DCR foi purificada na resina de afinidade Glutathione Sepharose 4B (Amersham-Biosciences) e a proteína GST foi removida por digestão com a protease TEV (*tobacco etch virus*). Os produtos da digestão foram incubados por diversas vezes na mesma resina até a remoção completa da GST e obtenção da

proteína DCR em bom grau de pureza (Figura 3A, artigo II). A massa molecular predita para esta proteína é 16967,5 Da, entretanto foi observado um padrão anormal de migração em gel de SDS-PAGE sendo o peso molecular maior do que o esperado (Figura 3A, artigo II). Um padrão eletroforético anormal também foi observado por Alves *et al.* (2004) e Alves & Castilho (2005) no estudo de uma proteína de *S. cerevisiae* nativamente desenhovelada.

A proteína DCR é rica em resíduos acídicos (glutamato e aspartato), representando 18% de sua composição, podendo resultar em baixa ligação do SDS levando à desnaturação incompleta e, conseqüentemente, migração lenta em gel de SDS-PAGE. Além disso, a assimetria desta região resultado da estrutura pouco enovelada que assume, pode produzir resistência anormal no gel de poliacrilamida. Muitas técnicas de biofísica têm sido empregadas para demonstrar a natureza de proteínas ou regiões de proteínas intrinsicamente desordenadas. As técnicas mais comuns usadas são dicroísmo circular, fluorescência e proteólise, embora o uso de técnicas como espalhamento de raio X a baixo ângulo e ressonância magnética nuclear esteja aumentando recentemente (Fontana *et al.*, 1997, Iakoucheva *et al.*, 2001, Fink, 2005).

Com o objetivo de determinar a correta massa da proteína DCR purificada, foi realizada uma análise por espectrometria de massas desta região. Foram obtidos os valores de 16965,25, 17041,25 e 17119,25 Da. Visto que esta região da proteína apresenta duas cisteínas, a diferença entre estes valores (76 Da) corresponde, provavelmente, a moléculas de β -mercaptoetanol ligadas a uma ou duas moléculas deste aminoácido, respectivamente (Figura 3B, artigo II).

Visando determinar se a proteína é desenhovelada confirmando os dados da predição de estrutura secundária, foram realizados ensaios de proteólise limitada, dicroísmo circular e fluorescência. Para o ensaio de proteólise limitada, 40 μ g de proteína foram digeridos quase completamente com tripsina (0,4 μ g) em apenas 10 minutos de digestão. Após 1 hora de incubação, a proteólise foi completa (Figura 3C, artigo II). A proteína DCR se mostrou muito sensível à ação da tripsina. Isto fica mais evidente quando comparada ao domínio SCAN submetido às mesmas condições de proteólise (Figura 3D, artigo II) e usado como controle, uma vez que os dados de

dicroísmo circular apresentados indicaram que ele é altamente estruturado (Figura 5A, artigo II).

O espectro de CD apresentou um mínimo bem evidente em 199 nm (Figura 4A, artigo II), compatível com proteínas intrinsecamente desordenadas que apresentam espectros de CD com sinal mais intenso na região de 200 nm e baixa elipticidade molar na região de 220. O espectro da análise de fluorescência apresentou um máximo em cerca de 354 nm (Figura 4B, artigo II). Os dados obtidos neste trabalho representam o somatório das emissões de fluorescência das duas moléculas de triptofano presentes na proteína. Os triptofanos livres em água ou em uma cadeia polipeptídica desestruturada apresentam emissão de fluorescência em torno de 350 a 360 nm. Estes dados corroboram a predição de estrutura secundária e os resultados obtidos com os ensaios de proteólise limitada e dicroísmo circular que indicam a natureza desordenada desta proteína.

Proteínas que apresentam regiões desordenadas estão freqüentemente envolvidas em funções regulatórias na célula como regulação da transcrição e tradução, transdução de sinal, fosforilação protéica, regulação da montagem de grandes complexos multiprotéicos como o flagelo bacteriano e o ribossomo (Dyson & Wright, 2005). Como a proteína NY-REN-21 é um possível fator de transcrição, a DCR pode ter função de facilitar a montagem do aparato de transcrição nas regiões promotoras dos genes, uma vez que é uma fração flexível da proteína além de poder mediar interações em escala variada de distância. Outra possibilidade é que esta região possa ser por si só uma região de interação entre proteínas.

4.3.3- Análise da região de dedos de zinco (*ZFR- Zinc Finger Region*)

É predito por análise da estrutura primária da proteína NY-REN-21 a presença de 7 domínios dedos de zinco na região C-terminal. Com o objetivo de estudar a estrutura secundária desta região, foram construídos os vetores pETGST-ZFR e pPROEX-ZFR, compreendendo os aminoácidos 275 a 473 da proteína NY-REN-21, para expressar as proteínas recombinantes GST-ZFR e His-ZFR, respectivamente, em *E. coli*.

Foram realizados testes de expressão na cepa de *E. coli* BL21(DE3) Δ slyD transformada com ambos os vetores nas temperaturas de indução de 20°C (indução por 16 horas com 10 mM de lactose), 25°C e 37°C (4 horas de indução, 0,5 mM de IPTG) e a fração solúvel dos extratos de *E. coli* foi analisada em gel de SDS-PAGE. Os melhores resultados foram obtidos com a proteína de fusão GST-ZFR, que apresentou boa quantidade na fração solúvel do extrato nas 3 temperaturas testadas (dados não mostrados). A proteína His-ZFR expressou de forma muito insolúvel com rendimento muito baixo na fração solúvel do extrato. A proteína de fusão GST-ZFR foi obtida em bom grau de pureza com o uso da resina Glutathione Sepharose (Figura 3A, artigo II), entretanto se mostrou muito instável, sendo totalmente degradada nos ensaios de digestão com a protease TEV (dados não mostrados). Além disso, esta proteína co-purifica com muito DNA e o tratamento com DNaseI, usado com sucesso para a purificação da proteína NY-REN-21, não foi eficaz neste caso. Devido à insolubilidade da proteína His-ZFR e da instabilidade da proteína de fusão GST-ZFR, não foi possível concluirmos as análises desta região.

4.3.4- SCAN: Purificação e análise por dicroísmo circular

A purificação do domínio SCAN foi realizada inicialmente por cromatografia de afinidade e apresentou grande rendimento e alto grau de pureza já nesta etapa da purificação. Posteriormente, a proteína foi digerida com a protease TEV para a remoção da cauda de histidina. Esta remoção foi necessária para melhor análise da

proteína por CD uma vez que este domínio apresenta cerca de 15 kDa sendo a porção da cauda de histidina juntamente com o sítio de clivagem desenovelados e podendo interferir no resultado do CD. Após a clivagem foi necessária uma cromatografia de troca iônica para a separação dos fragmentos digeridos, assim como da fração da proteína que não foi clivada. Conseguimos separar com sucesso a proteína digerida e obtê-la bastante pura para os ensaios de dicroísmo circular.

O domínio SCAN apresentou espectro característico de proteínas ricas em α -hélices, com dois mínimos bem definidos nos comprimentos de ondas de 208 e 222 nm e um máximo em torno de 192 nm (Figura 5, artigo II). Este resultado está de acordo com os dados obtidos por Stone *et al.*, 2002, que analisando a estrutura secundária do domínio SCAN da proteína ZNF174 encontrou cerca de 42% de estrutura tipo α -hélice. Os ensaios de desnaturação térmica foram realizados para observar a transição da proteína da forma estruturada para a desestruturada. Da mesma forma que para a proteína AtINT6 e suas formas truncadas, os dados de leitura a 222 nm foram normalizados para uma escala entre 0 e 1. Foram testadas duas concentrações (5 e 10 μ M) com o intuito de se observar se a proteína mais concentrada apresentaria T_m mais elevada, evidenciando que este domínio poderia estar dimerizado. O domínio SCAN se mostrou estável apresentado T_m de 60°C na concentração de 5 μ M e 62,8°C em 10 μ M (Figura 5B, artigo II). Este resultado foi similar ao encontrado para o domínio SCAN da proteína ZNF174, descrito como dímero, onde a variação da T_m dependente da concentração também foi observada, sendo de 75°C para a proteína 5 μ M e 72°C para concentração de 2 μ M (Stone *et al.*, 2002). A diferença de T_m dependente da concentração sugere o modelo de dois estados no qual o dímero enovelado está em equilíbrio com o monômero desenovelado, embora outros experimentos sejam necessários para comprovar que o domínio SCAN da proteína NY-REN-21 siga este modelo.

4.3.5- O domínio SCAN da proteína NY-REN-21 pode se associar em homodímeros e formar heterodímeros com a proteína SCAND1

O domínio SCAN tem sido relatado como uma região de dimerização, mediando a formação de homodímeros ou heterodímeros entre as proteínas que apresentam este domínio (Williams *et al.*, 1999; Sander *et al.*, 2000; Schumacher *et al.*, 2000). Williams *et al.* (1999) relatou a interação de diversas proteínas desta família através do método do duplo-híbrido em células de mamíferos. Neste estudo, a proteína ZFP38, possível ortóloga de camundongo da NY-REN-21, não formou homodímeros, se associando apenas a SRE-ZBP (*Serum Response Element- Zinc Finger Binding Protein*). No presente estudo, foi demonstrado por 3 técnicas diferentes a homodimerização do domínio SCAN da NY-REN-21.

Inicialmente, o cDNA correspondente ao domínio SCAN foi clonado nos vetores de duplo-híbrido de levedura pTL-1 (Carneiro *et al.*, 2006) e pGAD-C2 (James *et al.*, 1996), para testar a interação do domínio. A interação foi detectada pelos genes repórteres *HIS3* e *lacZ* do sistema de duplo-híbrido de levedura. A cepa contendo o domínio clonado nos dois vetores foi capaz de crescer em meio mínimo YNB (*Yeast Nitrogen Base*) contendo adenina e 15 mM de 3-AT (Figura 6A, artigo II). O ensaio de β -galactosidase também comprovou a interação (Figura 6B, artigo II).

A cromatografia por gel filtração foi realizada para estimar o peso molecular do domínio SCAN. Com esta finalidade o tempo de retenção da proteína de interesse na coluna foi comparado com proteínas padrões de peso molecular conhecido nas mesmas condições cromatográficas. O domínio SCAN apresentou apenas um pico de eluição com tempo de retenção de 77,40 minutos (dados não mostrados). Este tempo de retenção é intermediário entre os padrões quimiotripsinogênio 25 kDa (tempo de retenção: 85,87 minutos) e ovalbumina 43 kDa (tempo de retenção: 73,12 minutos), indicando que o tamanho do domínio SCAN está entre 25 e 43 kDa, compatível com o tamanho do dímero.

Além disso, o domínio SCAN recombinante também foi analisado para a determinação de sua massa e estado de oligomerização por espectrometria de massas. Esta análise da massa do domínio SCAN por *electrospray* confirmou a

formação de homodímeros (Figura 6C, artigo II). Para eliminar a possibilidade de que o dímero pudesse ter sido formado por ponte dissulfeto através da cisteína presente na seqüência, a amostra foi novamente analisada por espectrometria de massas na presença de 50% de acetonitrila e 0,1% de ácido fórmico. Nestas condições, onde os dímeros de cisteína não se desfazem, foi observado apenas o monômero (Figura 6D, artigo II), o que mostra que o homodímero foi formado por interações específicas não covalentes.

Com o objetivo de identificar outras proteínas capazes de dimerizar com o domínio SCAN da proteína NY-REN-21, nós realizamos uma triagem pelo método do duplo híbrido de levedura com uma biblioteca de cérebro fetal humano usando o domínio SCAN em fusão ao domínio de ligação ao DNA *lexA* de *E. coli* como isca. Foram seqüenciados 31 clones. O cDNA isolado com mais freqüência (5 cDNAs diferentes) codifica a proteína SCAND1 (*SCAN Domain containing 1*, também chamada SDP1 ou RAZ 1).

Esta interação foi confirmada pela retransformação da cepa de levedura L40 com os vetores pTL-SCAN e pACT-SCAND1. Como controle positivo foi usado o próprio domínio SCAN em fusão tanto ao domínio de ligação ao DNA *lexA* quanto ao domínio de ativação da transcrição *GAL4*. Para os controles negativos, foi utilizada a cepa L40 co-transformada com os pares de vetores pBTM-Nop8/pGAD-SCAN e pTL-1/pACT-SCAND1. As cepas resultantes foram crescidas em meio de cultura YNB acrescido de adenina e duas concentrações de 3-AT (15 e 30 mM). Nestas condições apenas as cepas co-transformadas com os vetores pTL-SCAN/pGAD-SCAN e pTL-SCAN/pACT-SCAND1 cresceram (Figura 7A, artigo II). As mesmas cepas foram usadas no ensaio de β -galactosidase que também confirmou a interação (Figura 7B, artigo II). As interações do domínio SCAN da proteína NY-REN-21, assim como da proteína NY-REN-21 completa com a proteína SCAND1 foram testadas através de um ensaio de co-expressão em *E. coli*. Neste ensaio, a proteína SCAND1 em fusão com o epítipo HA foi co-expressa com as proteínas recombinantes His-SCAN, His-NY-REN-21 e His-Nip7p (controle negativo) na cepa de *E. coli* BL21(DE3) Δ *slsD*. Os heterodímeros foram imobilizados na resina, lavados para remoção de ligantes não específicas e eluídos em 200 mM de imidazol. Estas

frações foram submetidas a *immunoblot* usando anticorpo contra o epítopo HA (Figura 7C, artigo II). Este ensaio confirmou a interação entre o antígeno NY-REN-21 e a proteína SCAND1, ressaltando que apenas o domínio SCAN da NY-REN-21 é necessário para a interação.

A proteína SCAND1 e sua ortóloga de camundongos PGC-2 (*PPAR gamma-coactivator 2*, Babb & Bowen, 2003), pertencem à família de proteínas que apresenta domínio SCAN, entretanto são particulares porque não possuem domínios de dedos de zinco associados. A proteína SCAND1 é expressa em diversos tecidos (Schumacher *et al.*, 2000; Sander & Morris, 2002), sendo os maiores níveis de expressão encontrados na próstata, rins, glândula tireóide, fígado e timo (Sander & Morris, 2002). A proteína também é expressa em linhagens celulares cancerígenas apresentando níveis de expressão mais elevados em células de carcinoma cervical e coloretal e adenocarcinoma (Sander & Morris, 2002).

O domínio SCAN é capaz de dimerizar seletivamente de forma que o estado de oligomerização deste domínio pode influenciar a atividade dos fatores de transcrição dedos de zinco (Schumacher *et al.*, 2000). A proteína SCAND1 pode apresentar grande importância nos mecanismos de regulação da transcrição. Além do antígeno NY-REN-21, foi descrito que a proteína SCAND1 interage também com os fatores de transcrição MZF1B (*Myeloid Zinc Finger 1B*, Sander *et al.*, 2000), ZNF 202, ZNF191 (Schumacher *et al.*, 2000) e com os receptores nucleares PPAR γ 2, α e δ e de estrogênio α (ER α , Babb & Bowen, 2003). Além disso, foi mostrado que SCAND1 foi capaz de romper a ligação do repressor transcrricional KAP1 de ZNF 202 *in vitro* (Porsh-Ozcurumez *et al.*, 2001). Sua interação com PPAR γ 2 aumentou a ativação da transcrição de PPAR γ 2 em células transientemente transfectadas (Babb & Bowen, 2003). Curiosamente, fusões das proteínas SCAND1 e PGC-2 ao domínio de ligação ao DNA GAL4 geraram proteínas incapazes de ativar ou reprimir a expressão do gene repórter. Entretanto, o mutante contendo apenas o domínio N-terminal, sem o domínio SCAN, apresentou propriedades de ativação da transcrição, sugerindo que o domínio SCAN de SCAND1 mascara um domínio de ativação transcrricional de SCAND1 (Babb & Bowen, 2003).

Existem poucos dados na literatura a respeito da proteína NY-REN-21 e seus parceiros de interação. Neste trabalho nós mostramos que o domínio SCAN da NY-REN-21 é capaz de se associar e interagir com a proteína SCAND1. O mapeamento das interações entre proteínas da família SCAN é importante, visto que variadas associações podem representar mecanismos diferentes de atuação dos fatores de transcrição.

5-CONCLUSÕES

-A proteína humana INT6 e seu domínio PCI quando em fusão ao domínio *lexA* de ligação ao DNA ativam os genes repórteres do sistema de duplo-híbrido de levedura da cepa L40, não sendo adequados para o uso desta metodologia.

-A proteína de fusão *lexA*-hINT6T22 não ativou os genes *HIS3* e *lacZ* da cepa L40, entretanto a triagem realizada para a identificação dos ligantes protéicos pelo método do duplo-híbrido de levedura apenas identificou proteínas falso-positivas. Provavelmente a fusão *lexA*-hINT6T22 não é funcional ou não interage com outras proteínas

-A proteína humana INT6 fusionada à cauda de histidina, proteína A e GST apresenta expressão muito baixa na fração solúvel do extrato de *E. coli*, ao contrário de sua ortóloga de *A. thaliana* que em fusão com cauda de histidina se apresenta parcialmente solúvel quando expressa a 25°C.

- O uso combinado dos métodos de proteólise limitada e espectrometria de massas permitiu a construção de 3 formas truncadas da proteína AtINT6, sendo que a forma truncada 3 (aminoácidos 172 a 415), que inclui o domínio PCI, apresenta alto nível de expressão, solubilidade e estabilidade, formando um domínio estrutural.

- A proteína NY-REN-21 é parcialmente solúvel quando expressa em *E. coli*, sendo a expressão a 20°C por 6 horas o protocolo de maior rendimento para a obtenção da proteína na fração solúvel do extrato. Sua purificação depende de tratamento com DNase e cromatografia em heparina para remover DNA e outras proteínas.

- A proteína NY-REN-21 é parcialmente desenovelada conforme análises de dicroísmo circular e sua estrutura secundária é sensível à adição de EDTA.

- Ensaios de proteólise limitada, dicroísmo circular e fluorescência revelaram que a região central da proteína NY-REN-21 que compreende os aminoácidos 128 a 276 é intrinsecamente desordenada.

- O domínio SCAN da proteína NY-REN-21 é rico em α -hélice e bastante estável de acordo com os ensaios de desnaturação térmica.

- O domínio SCAN da NY-REN-21 é capaz de formar homodímeros e heterodímeros com outra proteína da mesma família denominada SCAND1.

6- PERSPECTIVAS

Neste trabalho identificamos um fragmento da proteína INT6 de *A. thaliana* (aminoácidos 172 a 415) resistente à proteólise limitada que se mostrou altamente solúvel nos extratos de *E. coli* e estável após purificação. A identificação deste fragmento é bastante relevante visto que compreende o domínio PCI e pode ser usado para a determinação de sua estrutura tridimensional. O domínio PCI tem sido descrito como um domínio de interação entre proteínas (Yen & Chang, 2000; Kim *et al.*, 2001; Yahalom *et al.*, 2001) e não existe até o momento, informações de como pode reconhecer diferentes ligantes protéicos.

Em relação à proteína NY-REN-21, identificamos sua interação com a proteína SCAND1, entretanto a importância biológica desta interação não foi esclarecida. A proteína NY-REN-21 pode apresentar distintos mecanismos de ação, visto que pode formar homodímeros e heterodímeros com a proteína SCAND1. Estes mecanismos provavelmente estão relacionados ao controle da transcrição, uma vez que a proteína NY-REN-21 é um fator de iniciação da transcrição putativo. Ensaios em cultura de células de mamíferos para testar a atividade de transcrição da proteína NY-REN-21 e do heterodímero NY-REN-21-SCAND1 poderão comprovar esta hipótese.

Pelo menos 70 proteínas apresentam o domínio SCAN (Sander *et al.*, 2003). Este domínio é capaz de formar heterodímeros entre as proteínas que o contêm de forma específica. A determinação da estrutura tridimensional do heterodímero entre os domínios SCAN da proteína NY-REN-21 e SCAND1, assim como do homodímero poderá contribuir para o esclarecimento da especificidade das ligações.

A proteína NY-REN-21 apresenta 7 domínios dedos de zinco de interação com DNA e sua ortóloga em camundongo (proteína ZFP38) interage com uma sequência consenso 5'AGTAC3'. Estes fatos somados são um forte indício de que a proteína NY-REN-21 é capaz de ligar ao DNA de forma específica. A identificação da sequência de DNA a que esta proteína interage, assim como o isolamento de um possível gene alvo serão de grande importância para o entendimento do papel da proteína NY-REN-21 na célula.

7- REFERÊNCIAS BIBLIOGRÁFICAS

Ahmad, K. F., Melnick, A., Lax, S., Bouchard, D., Liu, J., Kiang, C. L., Mayer, S., Takahashi, S., Licht, J. D., Prive, G. G. 2003. Mechanism of SMRT corepressor recruitment by the BCL6 BTB domain. *Mol Cell*, 12:1551-1564.

Akiyoshi, Y., Clayton, J., Phan, L., Yamamoto, M., Hinnebusch, A. G., Watanabe, Y., Asano, K. 2001. Fission yeast homolog of murine Int-6 protein, encoded by mouse mammary tumor virus integration site, is associated with the conserved core subunits of eukaryotic translation initiation factor 3. *J Biol Chem*, 276:10056-10062.

Albagli-Curiel, O. 2003. Ambivalent role of BCL-6 in cell survival and transformation. *Oncogene*, 22:507-516.

Albright, S. R. & Tjian, R. 2000. TAFs revisited: more data reveal new twists and confirm old ideas. *Gene*, 242:1-13.

Alves, V. S. & Castilho, B. A. 2005. Gir2 is an intrinsically unstructured protein that is present in *Saccharomyces cerevisiae* as a group of heterogeneously electrophoretic migrating forms. *Biochem Biophys Res Commun*, 332:450-455.

Alves, V. S., Pimenta, D. C., Sattlegger, E., Castilho, B. A. 2004. Biophysical characterization of Gir2, a highly acidic protein of *Saccharomyces cerevisiae* with anomalous electrophoretic behavior. *Biochem Biophys Res Commun*, 314:229-234.

Asano, K., Merrick, W. C., Hershey, J. W. B. 1997. The translation initiation factor eIF3-p48 subunit is encoded by int-6, a site of frequent integration by the mouse mammary tumor virus genome. *J Biol Chem*, 272:23477-23480.

Babb, R. & Bowen, B. R. 2003. SDP1 is a peroxisome-proliferator-activated receptor gamma 2 co-activator that binds through its SCAN domain. *Biochem J*, 370:719-727.

Bachmann, F., Banziger, R., Burger, M. M. 1997. Cloning of a novel protein overexpressed in human mammary carcinoma. *Cancer Res*, 57:988-994.

Bandyopadhyay, A., Matsumoto, T., Maitra, U. 2000. Fission yeast Int6 is not essential for global translation initiation, but deletion of *int6+* causes hypersensitivity to caffeine and affects spore formation. *Mol Biol Cell*, 11:4005-4018.

Bardwell, V. J. & Treisman R. 1994. The POZ domain: a conserved protein-protein interaction motif. *Genes Dev*, 8:1664-1677.

Beato, M. & Klug, J. 2000. Steroid hormone receptors: an update. *Hum Reprod Update*, 6:225-236.

- Bellefroid, E. J., Poncelet, D. A., Lecocq, P. J., Revelant, O., Martial, J. A. 1991. The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc Natl Acad Sci U S A*, 88:3608-3612.
- Bomont, P., Cavalier, L., Blondeau, F., Ben Hamida, C., Belal, S., Tazir, M., Demir, E., Topaloglu, H., Korinthenberg, R., Tuysuz, B., Landrieu, P., Hentati, F., Koenig, M. 2000. The gene encoding gigaxonin, a new member of the cytoskeletal BTB/kelch repeat family, is mutated in giant axonal neuropathy. *Nat Genet*, 26:370-374.
- Borden, K. L. 2000. RING domains: master builders of molecular scaffolds? *J Moll Biol*, 295:1103-1112.
- Brown, T. A. 2002. Assembly of the Transcription Initiation Complex. In: *Genomes 2nd ed*. Taylor & Francis Group, London, UK.
- Browning, K. S., Gallie, D. R., Hershey, J. W., Hinnebuch, A. G., Maitra, U., Merrick, W. C., Norbury, C. 2001. Unified nomenclature for the subunits of eukaryotic initiation factor 3. *Trends Biochem Sci*, 26:284.
- Burley, S. K. & Roeder, R. G. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu Rev Biochem*, 65:769-799.
- Butler, J. E. & Kadonaga J. T. 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev*, 16:2583-2592.
- Buttitta, F., Martella, C., Barassi, F., Felicioni, L., Salvatore, S., Rosini, S., D'Antuono, T., Chella, A., Mucilli, F., Sacco, R., Mezzetti, A., Cuccurullo, F., Callahan, R., Marchetti, A. 2005. Int6 expression can predict survival in early-stage non-small cell lung cancer patients. *Clin Cancer Res*, 11:3198-3204.
- Carneiro, F. R., Silva, T. C., Alves, A. C., Haline-Vaz, T., Gozzo, F. C., Zanchin, N. I. 2006. Spectroscopic characterization of the tumor antigen NY-REN-21 and identification of heterodimer formation with SCAND1. *Biochem Biophys Res Commun*, 343:260-268.
- Castillo, G., Brun, R. P., Rosenfield, J. K., Hauser, S., Park, C. W., Troy, A. E., Wright, M. E., Spiegelman, B. M. 1999. An adipogenic cofactor bound by the differentiation domain of PPAR gamma. *EMBO J*, 18:3676-3687.
- Chen, G. & Burger, M. M. 1999. p150 expression and its prognostic value in squamous-cell carcinoma of the esophagus. *Int J Cancer*, 84:95-100.
- Chen, W., Cooper, T. K., Zahnow, C. A., Overholtzer, M., Zhou, Z., Ladanyi, M., Karp, J. E., Gokgoz, N., Wunder, J. S., Andrulis, I. L., Levine, A. J., Mankowski, J. L., Baylin, S. B. 2004. Epigenetic and genetic loss of Hic function accentuates the role of p53 in tumorigenesis. *Cancer Cell*, 6:387-398.

- Choo, Y. & Klug, A. 1993. A role in DNA binding for the linker sequences of the first three zinc fingers of TFIIIA. *Nucleic Acids Res*, 21:3341-3346.
- Chowdhury, K., Goulding, M., Walther., C., Imai, K., Fickenscher, H. 1992. The ubiquitous transactivator Zfp-38 is upregulated during spermatogenesis with differential transcription. *Mech Dev*, 39:129-142.
- Chung, H. R., Schafer, U., Jackle, H., Bohm, S. 2002. Genomic expansion and clustering of ZAD-containing C2H2 zinc-finger genes in Drosophila. *EMBO Rep*, 3:1158-1162.
- Coin, F. & Egly, J. M. 1998. Ten years of TFIIH. *Cold Spring Harb Symp Quant Biol*, 63:105-110.
- Collins, T., Stone, J. R., Williams, A. J. 2001. All in the family: the BTB/POZ, KRAB, and SCAN domains. *Mol Cell Biol*, 21:3609-3615.
- Cooper, G. M. 2000. RNA Synthesis and Processing. In: *The Cell - A Molecular Approach. 2nd ed.* Sinauer Associates, Inc. Sunderland, MA.
- Costoya, J. A. & Pandolfi, P. P. 2001. The role of promyelocytic leukemia zinc finger and promyelocytic leukemia in leukemogenesis and development. *Curr Opin Hematol*, 8:212-217.
- Crane, R., Craig, R., Murray, R., Dunand-Sauthier, I., Humphrey, T., Norbury, C. 2000. A fission yeast homolog of Int-6, the mammalian oncoprotein and eIF3 subunit, induces drug resistance when overexpressed. *Mol Biol Cell*, 10:3993-4003.
- Dahl, E., Koseki, H., Balling, R. 1997. Pax genes and organogenesis. *Bioessays*, 19:755-765.
- Dawid, I. B, Breen, J. J., Toyama, R. 1998. LIM domains: multiple roles as adapters and functional modifiers in protein interactions. *Trends Genet*, 14:156-162.
- Dellas, A., Torhorst, J., Bachmann, F., Banziger, R., Schultheiss, E., Burger, M. M. 1998. Expression of p150 in cervical neoplasia and its potencial value in predicting survival. *Cancer*, 83:1376-1383.
- deMartino, G. N. & Slaughter, C. A. 1999. The proteasome, a novel protease regulated by multiple mechanisms. *J Biol Chem*, 274:22123-22126.
- Desbois, C., Rousset, R., Bantignies, F., Jalinot, P. 1996. Exclusion of Int-6 from the PML nuclear bodies by binding to the HTLV-I Tax oncoprotein. *Science*, 273:951-953.
- Deuschle, U., Meyer, W. K., Thiesen, H. J. 1995. Tetracycline-reversible silencing of eukaryotic promoters. *Moll Cell Biol*, 15:1907-1914.

- Diella, F., Levi, G., Callahan, R. 1997. Characterization of the INT6 mammary tumor gene product. *DNA Cell Biol*, 16:839-847.
- Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., Obradovic, Z. 2001. Intrinsically disordered protein. *J Mol Graph Model*, 19:26-59.
- Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., Obradovic, Z. 2002. Intrinsic disorder and protein function. *Biochemistry*, 41:6573-6582.
- Dvir, A., Conaway, J. W., Conaway, R. C. 2001. Mechanism of transcription initiation and promoter escape by RNA polymerase II. *Curr Opin Genet Dev*, 11:209-214.
- Dyson, H. J. & Wright P. E. 2005. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*, 6:197-208.
- Eichmuller, S., Usener, D., Dummer, R., Stein, A., Thiel, D., Schadendorf, D. 2001. Serological detection of cutaneous T-cell lymphoma-associated antigens. *Proc Natl Acad Sci U S A*, 98:629-634.
- Fields, S. 2005. High-throughput two-hybrid analysis. The promise and the peril. *FEBS J*, 272:5391-5399.
- Fields, S., & Song, O. 1989. A novel genetic system to detect protein-protein interaction. *Nature*, 340:245-246.
- Fink, A.L. Natively unfolded proteins. 2005. *Curr Opin Struct Biol*, 15:35-41.
- Fontana, A., Zambonin, M., Polverino de Laureto, P., De Filippis, V., Clementi, A., Scaramella, E. 1997. Probing the conformational state of apomyoglobin by limited proteolysis. *J Mol Biol*, 266:223-230.
- Freilich, S., Oron, E., Kapp, Y., Nevo-Caspi, Y., Orgad, S., Segal, D., Chamovitz, D. A. 1999. The COP9 signalosome is essential for development of *Drosophila melanogaster*. *Curr Biol*, 9:1187-1190.
- Friedman, J. R., Fredericks, W. J., Jensen, D. E., Speicher, D. W., Huang, X. P., Neilson, E. G., Rauscher, F.J. 3rd. 1996. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev*, 10:2067-2078.
- Ganss, B. & Jheon, A. 2004. Zinc finger transcription factors in skeletal development. *Crit Rev Oral Biol Med*, 15:282-297.

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., Wuthrich, K. 1994. Homeodomain-DNA recognition. *Cell*, 78:211-223.

Geoffrey M. Cooper, 2000

Gimona, M., Djinovic-Carugo, K., Kranewitter, W. J., Winder, S. J. 2002. Functional plasticity of CH domains. *FEBS Lett*, 513:98-106.

Glickman, M. H., Rubin, D. M., Coux, O., Wefes, I., Pfeifer, G., Cjeda, Z., Baumeister, W., Fried, V. A., Finley, D. 1998. A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-Signalsome and eIF3. *Cell*, 94:615-623.

Green, M. R. 2000. TBP-associated factors (TAFII)s: multiple, selective transcriptional mediators in common complexes. *Trends Biochem Sci*, 25:59-63.

Guo, J. & Sen, G. C. 2000. Characterization of the interaction between the interferon-induced protein P56 and the Int-6 protein encoded by a locus of insertion of the mouse mammary tumor virus. *J Virol*, 74:1892-1899.

Hampsey, M. 1998. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev*, 62:465-503.

Hennemann, H., Vassen, L., Geisen, C., Eilers, M., Moroy, T. 2003. Identification of a novel Kruppel-associated box domain protein, Krim-1, that interacts with c-Myc and inhibits its oncogenic activity. *J Biol Chem*, 278:28799-28811.

Hershey, J. W. B. & Merrick, W. C. 2000. The pathway and mechanisms of initiation of protein synthesis. In: Sonenberg, N., Hershey, J. W. B., Mathews, M. B., eds. *Translational Control of Gene Expression*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Hihi, A. K., Michalik, L., Wahli, W. 2002. PPARs: transcriptional effectors of fatty acids and their derivatives. *Cell Mol Life Sci*, 59:790-798.

Hoareau Alves, K., Bochar, V., Rety S., Jalinot P. 2002. Association of the mammalian proto-oncoprotein Int-6 with the three protein complexes eIF3, COP9 signalsome and 26S proteasome. *FEBS Letters*, 527:15-21.

Hoatlin, M. E., Zhi, Y., Ball, H., Silvey, K., Melnick, A., Stone, S., Arai, S., Hawe, N., Owen, G., Zelent, A., Licht, J. D. 1999. A novel BTB/POZ transcriptional repressor protein interacts with the Fanconi anemia group C protein and PLZF. *Blood*, 94:3737-3747.

- Hofmann, K. & Bucher, P. 1998. The PCI domain: a common theme in three multiprotein complexes. *TIBS*, 23:204-205.
- Hollenberg, S. M., Sternglanz, R., Cheng, P. F. Weintraub, H. 1995. Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system. *Mol Cell Biol*, 15:3813-3822.
- Honer, C., Chen, P., Toth, M. J., Schumacher, C. 2001. The zinc finger protein 202 (ZNF202) is a transcriptional repressor of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene expression and a modulator of cellular lipid efflux. *J Biol Chem*, 276:12427-12433.
- Iakoucheva, L. M., Kimzey, A. L., Masselon, C. D., Bruce, J. E., Garner E. C., Brown, C. J., Dunker, A. K., Smith, R. D., Ackerman, E. J. 2001. Identification of intrinsic order and disorder in the DNA repair protein XPA. *Protein Sci*, 10:560-571.
- Ivanov, D., Stone, J. R., Maki, J. L., Collins, T., Wagner, G. 2005. Mammalian SCAN domain dimer is a domain-swapped homolog of the HIV capsid C-terminal domain. *Mol Cell*, 17:137-143
- James, P., Halladay, J., Craig, E. A. 1996. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics*, 144:1425-1436.
- Jones, D. T. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*, 292:195-202.
- Kamura, T., Koepp, D. M., Conrad, M. N., Skowyra, D., Moreland, R. J., Iliopoulos, O., Lane, W. S., Kaelin, W. G. Jr., Elledge, S. J., Conaway, R. C., Harper, J. W., Conaway, J. W. 1999. Rbx1, a component of the VHL tumor suppressor complex and SCF ubiquitin ligase. *Science*, 284:657-661.
- Kapelari, B., Bech-Otschir, D., Hegerl, R., Schade, R., Dumdey, R., Dubiel, W. 2000. Electron microscopy and subunit-subunit interaction studies reveal a first architecture of COP9 signalosome. *J Mol Biol*, 300:1169-1178.
- Karniol, B., Yahalom, A., Kwok, S., Tsuge, T., Matsui, M., Deng, X. W. 1998. The Arabidopsis homologue of an eIF3 complex subunit associates with the COP9 complex. *FEBS Lett*, 439:173-179.
- Kelly, S. M. & Price, N. C. 1997. The application of circular dichroism to studies of protein folding and unfolding. *Biochim Biophys Acta*, 338:161-185.
- Kerppola, T. & Curran, T. 1995. Transcription. Zen and the art of Fos and Jun. *Nature*, 373:199-200.

- Kim, T., Hofmann, K., von Armin, A. G., Chamovitz, D. A. 2001. PCI complexes: pretty complex interactions in diverse signaling pathways. *Trends Plant Sci*, 6:379-386.
- Kliwer, S. A., Lehmann, J. M., Milburn, M. V., Willson, T. M. 1999. The PPARs and PXR: nuclear xenobiotic receptors that define novel hormone signaling pathways. *Recent Prog Horm Res*, 54:345-367.
- Kobayashi, A., Yamagiwa, H., Hoshino, H., Muto, A., Sato, K., Morita, M., Hayashi, N., Yamamoto, M., Igarashi, K. 2000. A combinatorial code for gene expression generated by transcription factor Bach2 and MAZR (MAZ-related factor) through the BTB/POZ domain. *Mol Cell Biol*, 20:1733-1746.
- Kobayashi, A., Kang, M. I., Okawa, H., Ohtsuji, M., Zenke, Y., Chiba, T., Igarashi, K., Yamamoto, M. 2004. Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of Nrf2. *Mol Cell Biol*, 24:7130-7139.
- Koonin, E. V., Senkevich, T. G., Chernos, VI. 1992. A family of DNA virus genes that consists of fused portions of unrelated cellular genes. *Trends Biochem Sci*, 17:213-214.
- Laity, J. H., Dyson, H.J., Wright, P. E. 2000a. Molecular basis for modulation of biological function by alternate splicing of the Wilms' tumor suppressor protein. *Proc Natl Acad Sci U S A*, 97:11932-11935.
- Laity, J. H., Dyson, H. J., Wright, P. E. 2000b. DNA-induced alpha-helix capping in conserved linker sequences is a determinant of binding affinity in Cys(2)-His(2) zinc fingers. *J Mol Biol*, 295:719-727.
- Laity, J. H., Lee, B. M., Wright, P. E. 2001. Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struct Biol*, 11:39-46.
- Lamb, P., McKnight, S. L. 1991. Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. *Trends Biochem Sci*, 16:417-422.
- Lander, E. S., Linton, L. M., Birren, B. *et al.* 2001. International Human Genome Sequencing Consortium Initial sequencing and analysis of the human genome. *Nature*, 409:860-921.
- Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., Wright, P. E. 1989. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science*, 245:635-637.
- Li, Z., Wang, D., Na, X., Schoen, S. R., Messing, E. M., Wu, G. 2003. The VHL protein recruits a novel KRAB-A domain protein to repress HIF-1alpha transcriptional activity. *EMBO J*, 22:1857-1867.

- Littlewood, T. D. & Evan, G. I. 1995. Transcription factors 2: helix-loop-helix. *Protein Profile*, 2:621-702.
- Looman, C., Abrink, M., Mark, C., Hellman, L. 2002. KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. *Mol Biol Evol*, 19:2118-2130.
- Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R., Sigler, P. B. 1991. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature*, 352:497-505.
- Maeda, T., Hobbs, R. M., Merghoub, T., Guerna, I., Zelet, A., Cordon-Cardo, C., Teruya-Feldstein, J., Pandolfi, P. P. 2005. Role of proto-oncogene Pokemon in cellular transformation and ARF repression. *Nature*, 433:278-285.
- Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., Evans, R. M. 1995. The nuclear receptor superfamily: the second decade. *Cell*, 83:835-839.
- Mannini, R., Riviuccio, V., D'Auria, S., Tanfani, F., Ausili, A., Facchiano, A., Pedone, C., Grimaldi, G. 2006. Structure/function of KRAB repression domains: structural properties of KRAB modules inferred from hydrodynamic, circular dichroism, and FTIR spectroscopic analyses. *Proteins*, 62:604-616.
- Marchetti, A., Buttitta, F., Miyazaki, S., Gallahan, D., Smith, G. H., Gallahan, R. 1995. Int-6, a highly conserved, widely expressed gene, is mutated by mouse mammary tumor virus in mammary preneoplasia. *J Virol*, 69:1932-1938.
- Margolin, J. F., Friedman, J. R., Meyer, W. K., Vissing, H., Thiesen, H. J., Rauscher, F. J. 3rd. 1994. Kruppel-associated boxes are potent transcriptional repression domains. *Proc Natl Acad Sci U S A*, 91:4509-4513.
- Matt, T., Martinez-Yamout, M. A., Dyson, H. J., Wright, P. E. 2004. The CBP/p300 TAZ1 domain in its native state is not a binding partner of MDM2. *Biochem J*, 381:685-691.
- Matsumoto-Taniura, N., Pirollet, F., Monroe, R., Gerace, L., Westendorf, J. M. 1996. Identification of novel M phase phosphoproteins by expression cloning. *Mol Biol Cell*, 7:1455-1469.
- Mayeur, G. L. & Hershey, J. W. 2002. Malignant transformation by the eukaryotic translation initiation factor 3 subunit p48 (eIF3e). *FEBS Lett*, 514:49-54.
- McCarty, A. S., Kleiger, G., Eisenberg, D., Smale, S. T. 2003. Selective dimerization of a C2H2 zinc finger subfamily. *Mol Cell*, 11:459-470.

- McGuffin, L. J., Bryson, K., Jones, D. T. 2000. The PSIPRED protein structure prediction server. *Bioinformatics*, 16: 404-405.
- Melnick, A., Ahmad, K. F., Arai, S., Polinger, A., Ball, H., Borden, K. L., Carlile, G. W., Prive, G. G., Licht, J. D. 2000. In-depth mutational analysis of the promyelocytic leukemia zinc finger BTB/POZ domain reveals motifs and residues required for biological and transcriptional functions. *Mol Cell Biol*, 20:6550-6567.
- Miyazaki, S., Imatani, A., Ballard, L., Marchetti, A., Buttitta, F., Albertsen, H., Nevanlinna, H. A., Gallahan, D., Callahan, R. 1997. The chromosome location of the human homolog of the mouse mammary tumor associated gene INT6 and its status in human breast carcinomas. *Genomics*, 46:155-158.
- Miyazaki, S., Rasmussen, S., Imatani, A., Diella, F., Sullivan, D. T., Callahan, R. 1999. Characterization of the Drosophila ortholog of mouse eIF-3p48/INT-6. *Gene (Amst)*, 233: 241-247.
- Moosmann, P., Georgiev, O., Le Douarin, B., Bourquin, J. P., Schaffner, W. 1996. Transcriptional repression by RING finger protein TIF1 beta that interacts with the KRAB repressor domain of KOX1. *Nucleic Acids Res*, 24:4859-4867.
- Moosmann, P., Georgiev, O., Thiesen, H. J., Hagmann, M., Schaffner, W. 1997. Silencing of RNA polymerases II and III-dependent transcription by the KRAB protein domain of KOX1, a Kruppel-type zinc finger factor. *Biol Chem*, 378:669-677.
- Morgan, B., Sun, L., Avitahl, N., Andrikopoulos, K., Ikeda, T., Gonzales, E., Wu, P., Neben, S., Georgopoulos, K. 1997. Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. *EMBO J*, 16:2004-2013.
- Morris-Desbois, C., Bochard, V., Reynaud, C., Jalinot, P. 1999. Interaction between the Ret finger protein and the Int-6 gene product and co-localisation into nuclear bodies. *J Cell Sci*, 112: 3331-3342.
- Morris-Desbois, C., Rety, S., Myriam F., Garin, J., Jalinot, P. 2001. The human protein HSPC021 interacts with Int-6 and is associated with Eukaryotic translation initiation factor 3. *J Biol Chem*, 276: 45988-45995.
- Nam, K., Honer, C., Schumacher, C. 2004. Structural components of SCAN-domain dimerizations. *Proteins*, 56:685-692.
- Oikawa, T. & Yamada, T. 2003. Molecular biology of the Ets family of transcription factors. *Gene*, 303:11-34.
- Orphanides, G., Lagrange, T., Reinberg, D. 1996. The general transcription factors of RNA polymerase II. *Genes Dev*, 10:2657-2683.

- Pavletich, N. P. & Pabo, C. O. 1993. Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science*, 261:1701-1707.
- Peng, H., Begg, G. E., Harper, S. L., Friedman, J. R., Speicher, D. W., Rauscher, F. J. 3rd. 2000a. Biochemical analysis of the Kruppel-associated box (KRAB) transcriptional repression domain. *J Biol Chem*, 275:18000-18010.
- Peng, H., Begg, G. E., Schultz, D. C., Friedman, J. R., Jensen, D. E., Speicher, D. W., Rauscher, F. J. 3rd. 2000b. Reconstitution of the KRAB-KAP-1 repressor complex: a model system for defining the molecular anatomy of RING-B box-coiled-coil domain-mediated protein-protein interactions. *J Mol Biol*, 295:1139-1162.
- Pengue, G., Calabro, V., Bartoli, P. C, Pagliuca, A., Lania, L. 1994. Repression of transcriptional activity at a distance by the evolutionarily conserved KRAB domain present in a subfamily of zinc finger proteins. *Nucleic Acids Res*, 22:2908-2914.
- Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilipenko, E. V., Shatsky, I. N., Agol, V. I., Hellen, C. U. 2001. *Proc Natl Acad Sci U S A*, 98:7029-7036.
- Peukert, K., Staller, P., Schneider, A., Carmichael, G., Hanel, F., Eilers, M. 1997. An alternative pathway for gene regulation by Myc. *EMBO J*, 16:5672-5686.
- Pincheira, R., Chen, Q., Zhang, J. T. 2001. Identification of a 170-kDa protein over-expressed in lung cancers. *Br J Cancer*, 84:1520-1527.
- Pintard, L., Willis, J. H., Willems, A., Johnson, J. L., Srayko, M., Kurz, T., Glaser, S., Mains, P. E., Tyers, M., Bowerman, B., Peter, M. 2003. The BTB protein MEL-26 is a substrate-specific adaptor of the CUL-3 ubiquitin-ligase. *Nature*, 425:311-316.
- Pinte, S., Stankovic-Valentin, N., Deltour, S., Rood, B. R., Guerardel, C., Leprince, D. 2004. The tumor suppressor gene HIC1 (hypermethylated in cancer 1) is a sequence-specific transcriptional repressor: definition of its consensus binding sequence and analysis of its DNA binding and repressive properties. *J Biol Chem*, 279:38313-38324.
- Polo, J. M., Dell'Oso, T., Ranuncolo, S. M., Cerchietti, L., Beck, D., Da Silva, G. F., Prive, G. G., Licht, J. D., Melnick, A. 2004. Specific peptide interference reveals BCL6 transcriptional and oncogenic mechanisms in B-cell lymphoma cells. *Nat Med*, 10:1329-1335.
- Poncelet, D. A., Bellefroid, E. J., Bastiaens, P. V., Demoitie, M. A., Marine, J. C., Pendeveille, H., Alami, Y., Devos, N., Lecocq, P., Ogawa, T., Muller, M., Martial, J. A. 1998. Functional analysis of ZNF85 KRAB zinc finger protein, a member of the highly homologous ZNF91 family. *DNA Cell Biol*, 17:931-943.
- Porsch-Ozcurumez, M., Langmann, T., Heimerl, S., Borsukova, H., Kaminski, W. E., Drobnik, W., Honer, C., Schumacher, C., Schmitz, G. 2001. The zinc finger protein

202 (ZNF202) is a transcriptional repressor of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene expression and a modulator of cellular lipid efflux. *J Biol Chem*, 276:12427-12433.

Rasmussen, S. B., Kordon, E., Callahan, R., Smith, G. H. 2001. Evidence for the transforming activity of a truncated int6 gene, *in vitro*. *Oncogene*, 20:5291-5301.

Rastinejad, F. 2001. Retinoid X receptor and its partners in the nuclear receptor family. *Curr Opin Struct Biol*, 11:33-8.

Roeder, R. G. 1991. The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends Biochem Sci*, 16:402-408.

Roeder, R. G. 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci*, 21:327-335.

Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., Dunker, A. K. 2001. Sequence complexity of disordered protein. *Proteins*, 42:38-48.

Rosati, M., Marino, M., Franze, A., Tramontano, A., Grimaldi, G. 1991. Members of the zinc finger protein gene family sharing a conserved N-terminal module. *Nucleic Acids Res*, 19:5661-5667.

Rosati, M., Franze, A., Matarazzo, M. R., Grimaldi, G. 1999. Coding region intron/exon organization, alternative splicing, and X-chromosome inactivation of the KRAB/FPB-domain-containing human zinc finger gene ZNF41. *Cytogenet Cell Genet*, 85:291-296.

Rosati, M., Rocchi, M., Storlazzi, C.T., Grimaldi, G. 2001. Assignment to chromosome 12q24.33, gene organization and splicing of the human KRAB/FPB containing zinc finger gene ZNF84. *Cytogenet Cell Genet*, 94:127-130.

Sander, T. L. & Morris, J. F. 2002. Characterization of the SCAN box encoding RAZ1 gene: analysis of cDNA transcripts, expression, and cellular localization. *Gene*, 296: 53-64.

Sander, T. L., Haas, A. L., Peterson, M. J., Morris, J. F. 2000. Identification of a novel SCAN box-related protein that interacts with MZF1B. The leucine-rich SCAN box mediates hetero- and homoprotein associations. *J Biol Chem*, 275:12857-12867.

Sander, T. L., Stringer, K. F, Maki, J. L., Szauter, P., Stone, J. R., Collins, T. 2003. The SCAN domain defines a large family of zinc finger transcription factors. *Gene*, 310:29-38.

Scanlan, M. J., Gordan, J. D., Williamson, B., Stockert, E., Bander, N. H., Jongeneel, V., Gure, A. O., Jager, D., Jager, E., Knuth, A., Chen, Y. T., Old, L. J. 1999. Antigens

recognized by autologous antibody in patients with renal-cell carcinoma. *Int J Cancer*, 83:456-464.

Scheel, H. & Hofmann, K. 2005. Prediction of a common structural scaffold for proteasome lid. COP9-signalosome and eIF3 complexes. *BMC Bioinformatics*, 6:71.

Schultz, J., Milpetz, F., Bork, P., Ponting, C. P. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A*, 95:5857-5864.

Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., Bork, P. 2000. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res*, 28:231-234.

Schumacher, C., Wang, H., Honer, C., Ding, W., Koehn, J., Lawrence, Q., Coulis, C. M., Wang, L. L., Ballinger, D., Bowen, B. R., Wagner, S. 2000. The SCAN domain mediates selective oligomerization. *J Biol Chem*, 275:17173-17179.

Schwechheimer, C. & Deng, X. W. 2001. COP9 signalosome revisited: a novel mediator of protein degradation. *Trends Cell Biol*, 11:420-426.

Senatore, B., Cafieri, A., Di Marino, I., Rosati, M., Di Nocera, P. P., Grimaldi, G. 1999. A variety of RNA polymerases II and III-dependent promoter classes is repressed by factors containing the Kruppel-associated/finger preceding box of zinc finger proteins. *Gene*, 234:381-394.

Shalev, A., Valasek, L., Pise-Masison, C. A., Radonovich, M., Phan, L., Clayton, J., He, H., Brady, J. N., Hinnebusch, A. G., Asano, K. 2001. *Saccharomyces cerevisiae* protein Pci8p and human protein eIF3e/Int-6 interact with the eIF3 core complex by binding to cognate eIF3b subunits. *J Biol Chem*, 276:34948-34957.

Shimono, Y., Murakami, H., Hasegawa, Y., Takahashi, M. 2000. RET finger protein is a transcriptional repressor and interacts with enhancer of polycomb that has dual transcriptional functions. *J Biol Chem*, 275:39411-39419.

Smale, S. T. & Kadonaga, J. T. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem*, 72:449-479.

Sonenberg, N. & Dever, T. E. 2003. Eukaryotic translation initiation factors and regulators.. *Curr Opin Struct Biol*, 13:56-63.

Sreerama, N., Manning, M. C., Poweres, M. E., Zhang, J. X., Goldenberg, D. P., Woody, R. W. 1999. Tyrosine, phenylalanine and disulfide contributions to the circular dichroism of proteins: circular dichroism spectra of wild-type and mutant bovine pancreatic trypsin inhibitor. *Biochemistry*, 38:10814-10822.

- Stogios, P. J., Downs, G. S., Jauhal, J. J., Nandra, S. K., Prive, G. G. 2005. Sequence and structural analysis of BTB domain proteins. *Genome Biol*, 6:R82.
- Stone, J. R., Maki, J. L., Blacklow, S. C., Collins, T. 2002. The SCAN domain of ZNF174 is a dimer. *J Biol Chem*, 277:5448-5452.
- Sun, L., Liu, A., Georgopoulos, K. 1996. Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development. *EMBO J*, 15:5358-5369.
- Sun, X., Perlick, H. A., Dietz, H. C., Maquat, L. E. 1998. A mutated human homologue to yeast Upf1 protein has a dominant-negative effect on the decay of nonsense-containing mRNAs in mammalian cells. *Proc Natl Acad Sci U S A*, 95:10009-10014.
- Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Suyama, A., Sakaki, Y., Morishita, S., Okubo, K., Sugano, S. 2001. Identification and Characterization of the Potential Promoter of Human Genes. *Genome Res*, 11:677-684.
- Takenaga, M., Hatano, M., Takamori, M., Yamashita, Y., Okada, S., Kuroda, Y., Tokuhisa, T. 2003. Bcl6-dependent transcriptional repression by BAZF. *Biochem Biophys Res Commun*, 303:600-608.
- Tjian, R. & Maniatis, T. 1994. Transcriptional activation: a complex puzzle with few easy pieces. *Cell*, 77:5-8.
- Topcu, Z., Mack, D. L., Hromas, R. A., Borden, K. L. 1999. The promyelocytic leukemia protein PML interacts with the proline-rich homeodomain protein PRH: A RING may link hematopoiesis and growth control. *Oncogene*, 18:7091-7100.
- Tsuge, T., Matsui, M., Wei, N. 2001. The subunit 1 of the COP9 signalosome suppresses gene expression through its N-terminal domain and incorporates into the complex through the PCI domain. *J Mol Biol*, 305:1-9.
- Uversky, V. N. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*, 11:739-756.
- Venter, J. C., Adams, M. D., Myers, E. W. *et al.* 2001. The sequence of the human genome. *Science*, 291:1304-1351.
- Vidal, M. & Legrain, P. 1999. Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res*, 27:919-929.
- Vinson, C., Myakishev, M., Acharya, A., Mir, A. A., Moll, J. R., Bonovich, M. 2002. Classification of human B-ZIP proteins based on dimerization properties. *Mol Cell Biol*, 22:6321-6335.

- Vissing, H., Meyer, W. K., Aagaard, L., Tommerup, N., Thiesen, H. J. 1995. Repression of transcriptional activity by heterologous KRAB domains present in zinc finger proteins. *FEBS letters*, 369:153-157.
- von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., Bork, P. 2002. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 417:399-403.
- Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., Jones, D. T. 2004. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol*, 337:635-645.
- Watkins, S. J., Norbury, C. J. 2004. Cell cycle-related variation in subcellular localization of eIF3e/INT6 in human fibroblasts. *Cell Prolif*, 37:149-160.
- Wei, N. & Deng, X. W. 1998. Characterization and purification of the mammalian COP9 complex, a conserved nuclear regulator initially identified as a repressor of photomorphogenesis in higher plants. *Photochem Photobiol*, 68:237-241.
- Westendorf, J. J. & Hiebert, S. W. 1999. Mammalian runt-domain proteins and their roles in hematopoiesis, osteogenesis, and leukemia. *J Cell Biochem*, 75:51-58.
- Wilkins, A., Ping, Q., Carpenter, C. L. 2004. RhoBTB2 is a substrate of the mammalian Cul3 ubiquitin ligase complex. *Genes Dev*, 18:856-861.
- Williams, A. J., Khachigian, L. M., Shows, T., Collins, T. 1995. Isolation and characterization of a novel zinc-finger protein with transcription repressor activity. *J Biol Chem*, 270:22143-22152.
- Williams, A. J., Blacklow, S. C., Collins, T. 1999. The zinc finger-associated SCAN box is a conserved oligomerization domain. *Mol Cell Biol*, 19:8526-8535.
- Wintjens, R. & Rooman, M. 1996. Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *J Mol Biol*, 262:294-313.
- Witzgall, R., O'Leary, E., Leaf, A., Onaldi, D., Bonventre, J. V. 1994. The Kruppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proc Natl Acad Sci U S A*, 91:4514-4518.
- Wolfe, S. A., Nekludova, L., Pabo, C. O. 2000. DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct*, 29:183-212.
- Woody, R. W. 1995. Circular dichroism. *Methods Enzymol*, 246:34-71.

- Xu, L., Wei, Y., Reboul, J., Vaglio, P., Shin, T. H., Vidal, M., Elledge, S. J., Harper J. W. 2003. BTB proteins are substrate-specific adaptors in an SCF-like modular ubiquitin ligase containing CUL-3. *Nature*, 425:316-321.
- Yahalom, A., Kim, T. H., Winter, E., Karniol, B., von Arnim, A. G., Chamovitz, D. A. 2001. Arabidopsis eIF3e (INT-6) associates with both eIF3c and the COP9 signalosome subunit CSN7. *J Biol Chem*, 276:334-340.
- Yang, X. W., Zhong, R., Heintz, N. 1996. Granule cell specification in the developing mouse brain as defined by expression of the zinc finger transcription factor RU49. *Development*, 122:555-566.
- Yang, X. W., Wynder, C., Doughty, M. L., Heintz, N. 1999. BAC-mediated gene-dosage analysis reveals a role for Zipro1 (Ru49/Zfp38) in progenitor cell proliferation in cerebellum and skin. *Nat Genet*, 22:327-335.
- Yen, H. C. S. & Chang, E. C. 2000. Yin6, a fission yeast Int6 homolog, complexes with Moe 1 and plays a role in chromosome segregation. *Cell Biology*, 97:14370-14375.
- Yen, H. C., Gordon, C., Chang, E. C. 2003. Schizosaccharomyces pombe Int6 and Ras homologs regulate cell division and mitotic fidelity via the proteasome. *Cell*, 112:207-217.
- Yoshida, M. Multiple viral strategies of HTLV-1 for disregulation of cell growth control. 2001. *Annu Rev Immunol*, 19:475-496.
- Zawel, L. & Reinberg, D. 1995. Common themes in assembly and function of eukaryotic transcription complexes. *Annu Rev Biochem*, 64:533-561.
- Ziegelbauer, J., Shan, B., Yager, D., Larabell, C., Hoffmann, B., Tjian, R. 2001. Transcription factor MIZ-1 is regulated via microtubule association. *Mol Cell*, 8:339-349.
- Zollman, S., Godt, D., Prive, G. G., Couderc, J. L., Laski, F. A. 1994. The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in Drosophila. *Proc Natl Acad Sci U S A*, 91:10717-10721.

8- ANEXOS

8.1- ANEXO I

Smetana, J. H., Oliveira, C. L., Jablonka, W., Aguiar Pertinhez, T., Carneiro, F. R., Montero-Lomeli, M., Torriani, I., Zanchin, N. I. 2006. Low resolution structure of the human alpha4 protein (IgBP1) and studies on the stability of alpha4 and of its yeast ortholog Tap42. *Biochim Biophys Acta*. Feb 17; [Epub ahead of print].

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Biochimica et Biophysica Acta xx (2006) xxx–xxx

BIOCHIMICA ET BIOPHYSICA ACTA
BBA<http://www.elsevier.com/locate/bba>

Low resolution structure of the human $\alpha 4$ protein (IgBP1) and studies on the stability of $\alpha 4$ and of its yeast ortholog Tap42

Juliana Helena Costa Smetana ^a, Cristiano Luiz Pinto Oliveira ^{a,b}, Willy Jablonka ^c,
Thelma Aguiar Pertinhez ^{a,1}, Flavia Raquel Gonçalves Carneiro ^a, Monica Montero-Lomeli ^c,
Iris Torriani ^{a,b}, Nilson Ivo Tonin Zanchin ^{a,*}

^a Brazilian Synchrotron Light Laboratory (LNLS), Centro de Biologia Molecular Estrutural, Laboratório Nacional de Luz Síncrotron,
R. Giuseppe Máximo Scolfaro, 10.000, Campinas - SP, PO Box 6192-CEP 13084-971, Brazil

^b Instituto de Física “Gleb Wataghin”, Unicamp, Campinas SP 13084-971, Brazil

^c Instituto de Bioquímica Médica, Programa de Biologia Molecular e Biotecnologia, Universidade Federal do Janeiro, Rio de Janeiro, RJ 21941-590, Brazil

Received 15 September 2005; received in revised form 15 January 2006; accepted 20 January 2006

Abstract

The yeast Tap42 and mammalian $\alpha 4$ proteins belong to a highly conserved family of regulators of the type 2A phosphatases, which participate in the rapamycin-sensitive signaling pathway, connecting nutrient availability to cell growth. The mechanism of regulation involves binding of Tap42 to Sit4 and PPH21/22 in yeast and binding of $\alpha 4$ to the catalytic subunits of type 2A-related phosphatases PP2A, PP4 and PP6 in mammals. Both recombinant proteins undergo partial proteolysis, generating stable N-terminal fragments. The full-length proteins and $\alpha 4$ C-terminal deletion mutants at amino acids 222 ($\alpha 4\Delta 222$), 236 ($\alpha 4\Delta 236$) and 254 ($\alpha 4\Delta 254$) were expressed in *E. coli*. $\alpha 4\Delta 254$ undergoes proteolysis, producing a fragment similar to the one generated by full-length $\alpha 4$, whereas $\alpha 4\Delta 222$ and $\alpha 4\Delta 236$ are highly stable proteins. $\alpha 4$ and Tap42 show α -helical circular dichroism spectra, as do their respective N-terminal proteolysis resistant products. The cloned truncated proteins $\alpha 4\Delta 222$ and $\alpha 4\Delta 236$, however, possess a higher content of α -helix, indicating that the C-terminal region is less structured, which is consistent with its higher sensitivity to proteolysis. In spite of their higher secondary structure content, $\alpha 4\Delta 222$ and $\alpha 4\Delta 236$ showed thermal unfolding kinetics similar to the full-length $\alpha 4$. Based on small angle X-ray scattering (SAXS), the calculated radius of gyration for $\alpha 4$ and Tap42 were 41.2 ± 0.8 Å and 42.8 ± 0.7 Å and their maximum dimension ~ 142 Å and ~ 147 Å, respectively. The radii of gyration for $\alpha 4\Delta 222$ and $\alpha 4\Delta 236$ were 21.6 ± 0.3 Å and 25.7 ± 0.2 Å, respectively. Kratky plots show that all studied proteins show variable degree of compactness. Calculation of model structures based on SAXS data showed that $\alpha 4\Delta 222$ and $\alpha 4\Delta 236$ proteins have globular conformation, whereas $\alpha 4$ and Tap42 exhibit elongated shapes.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Tap42 protein family; Rapamycin signaling pathway; Small angle X-ray scattering

1. Introduction

The control of cell growth in response to nutrients and growth factors is performed by a rapamycin sensitive pathway which acts on protein synthesis, transcription, nutrient transporter turnover and progression of the cell cycle. The major signaling protein in this pathway is Tor, the Target of Rapamycin, which

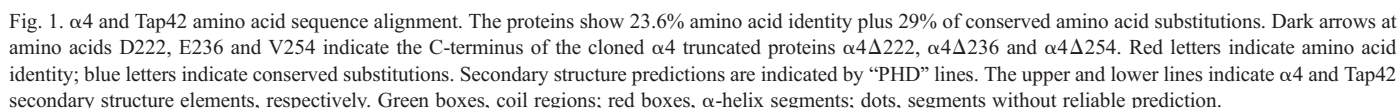
belongs to the family of phosphatidyl-inositol-kinase-related kinases [1,2]. In yeast cells, the control of protein synthesis by this pathway involves mainly the regulation of the type 2A-related phosphatases Sit4 and PPH22 by the regulatory protein Tap42 [3]. Under growth-promoting conditions, yeast Tor2 phosphorylates Tap42 and thus promotes its association with Sit4 and PPH22, which reduces their phosphatase activity and allows activation of the translation initiation factor 4E [1].

The yeast Tap42 and its mammalian ortholog $\alpha 4$, also known as Immunoglobulin Binding Protein 1 (IgBP1), share 52.6% amino acid similarity (Fig. 1) and belong to a unique and highly conserved family [4–7]. Mammalian $\alpha 4$ has been implicated in

* Corresponding author. Tel.: +55 19 3512 1113; fax: +55 19 3512 1004.

E-mail address: zanchin@lnls.br (N.I.T. Zanchin).

¹ Present address: University of Parma, Dept. of Experimental Medicine, 43100 Parma, Italy.



Association of $\alpha 4$ or the A/B subunits to PP2Ac is mutually exclusive due to an overlapping in the binding site and binding of $\alpha 4$ to PP2Ac causes the displacement of the A and B subunits [15]. PP2Ac distribution between these two complexes is regulated by post-translational modifications [17]. The interaction between murine $\alpha 4$ and PP2Ac has been mapped to residues 94–202 and 19–165, respectively [15,18]. A more refined mapping showed that residues 19–22 and 150–164 of PP2A are essential for binding to $\alpha 4$ [19]. The association between PP2Ac and $\alpha 4$ is generally thought to be part of a

Despite the large number of studies on the function of $\alpha 4$ and Tap42, the structural basis for their regulatory function is still unclear. This work presents the first structural data for proteins of the Tap42 family, obtained using standard biophysical techniques, and shows that they belong to the recently identified class of intrinsically unstructured proteins. Proteins of this type contain large unstructured regions, but they frequently undergo induced folding upon binding to a partner such as DNA, membranes or other proteins. Their flexibility is functionally important, since it allows them to fit a variety of different binding partners and also provides a larger binding interface when compared to folded proteins of the same size [23]. Intrinsically unstructured regions can be predicted from the primary structure and several algorithms have been developed for this purpose [24]. In general, these regions are characterized by a high net charge and a low

hydrophobicity [25]. A comparison of the predicted unstructured regions of proteins from the three domains of life has shown that they are far more frequent in eukaryotic genomes than in bacterial or archaeal genomes [26]. They are also more frequently found in regulatory and cancer-associated proteins than in those involved in metabolism, biosynthesis, and degradation [27]. Our findings provide a structural framework to understand the cellular function of $\alpha 4$ and Tap42 in the context of intrinsic disorder.

2. Materials and methods

2.1. Plasmid construction

The $\alpha 4$ cDNA was amplified by PCR from a human fetal brain cDNA library (Clontech) using primers ONZ107: 5' CCCCCATATGGCTGCTGAGGAC-GAGTTAC 3' (forward) and ONZ108: 5' TGGGATCCTCAGCC-CATGTTCTGTCGGTTC 3' (reverse). The PCR product was digested with *NdeI* and *BamHI* and cloned into the pET28a vector. cDNAs of truncated forms of $\alpha 4$ were generated by PCR using the full-length $\alpha 4$ cDNA as a template and subcloned into either pET28a (in frame with an N-terminal hexahistidine tag) or pET29a. All DNA vectors and inserts were digested with *NdeI* and *BamHI* prior to ligation. The following deletion mutants (Fig. 1) were constructed: pET28a- $\alpha 4\Delta 254$ (amino acids 1–254, using primers ONZ107 described above and ONZ167: 5' GGATCCTTAATACTTTGGCTTGAGCCATG 3'), pET28a- $\alpha 4\Delta 236$ (amino acids 1–236, using primers ONZ107 and ONZ191: 5' GGATCCTTACTCCTGGCGAGATGAGTTAGAAG 3'), pET28a- $\alpha 4\Delta 222$ (amino acids 1–222, using primers ONZ107 and ONZ192: 5' GGATCCT-TAGTCTCTTTCTCTCAGGATCTTTA 3'). Vector pET29a- $\alpha 4\Delta 222$ was constructed by transferring the cDNA sequence comprising nucleotides 1–675 from pET28a- $\alpha 4\Delta 222$ to pET29a. The *TAP42* gene was amplified by PCR from yeast genomic DNA using primers 5' GAGAATTCGGATCCGCGTCAGTAA-CAG 3' (forward) and 5' GAGCTGGATTCCGTACCCATCCTCTATTTC 3' (reverse) and cloned into the pGEM-T Easy Vector (Promega). Subsequently, the *TAP42* gene was digested with *KpnI* and *BamHI* and subcloned into the pQE30 plasmid (Qiagen) in frame with an N-terminal hexahistidine tag. Standard molecular biology techniques were performed as described elsewhere [28]. All of the constructs were verified by DNA sequencing analysis.

2.2. Recombinant protein expression

The human $\alpha 4$ and its deletion mutants were expressed in *E. coli* BL21 (DE3) carrying the plasmid pRARE (Novagen), which encodes tRNAs for rare codons in *E. coli*. Tap42 was expressed in *E. coli* M15 carrying the plasmid pREP4, which encodes for the T5 RNA polymerase repressor. Pre-cultures of transformed cells were grown overnight at 37 °C and 200 rpm in LB broth containing the appropriate antibiotics (30 μ g/mL ampicillin and 25 μ g/mL kanamycin for Tap42; 20 μ g/mL chloramphenicol and 50 μ g/mL kanamycin for $\alpha 4$ and its deletion mutants). The pre-cultures were used to inoculate fresh LB broth at 2%, which was incubated at 37 °C until the optical density at 600 nm reached ~ 1.0 . At this time, expression of the recombinant proteins was induced with 0.5 mM IPTG and the cultures were incubated for 4 h at 25 °C (for $\alpha 4$ and its deletion mutants) or 2 h at 37 °C (for Tap42). Cells were harvested by centrifugation at 4000 $\times$ g for 10 min at 4 °C and stored at –20 °C.

2.3. Purification of his-tagged proteins

Induced bacterial cells were suspended in affinity chromatography buffer (50 mM sodium phosphate buffer, pH 7.2, 100 mM NaCl, 5% glycerol, 5 mM imidazole, 1 mM PMSF) and incubated on ice with lysozyme (50 μ g/mL) for 30 to 60 min. Cells were disrupted by 5–8 cycles of sonication and the lysate was centrifuged at 30,000 $\times$ g for 30 min at 4 °C. The supernatants were passed by gravity through columns containing Ni-NTA Superflow resin (Qiagen) pre-equilibrated with affinity buffer. The columns were washed with affinity buffer and eluted with increasing imidazole concentrations (up to 200 mM) in the same buffer. The proteins were dialyzed against ion exchange buffer A (Tris–HCl

10 mM, pH 7.4, 10 mM NaCl, 5 mM EDTA, 1 mM benzamidine, 5 mM β -mercaptoethanol, 1 mM PMSF and 1 μ g/mL pepstatin A) and applied onto a 5-mL HiTrap Q-Sepharose HP column (Amersham Biosciences) pre-equilibrated with ion exchange buffer A. The column was washed with 10 mL buffer A and proteins were eluted with a 90-mL linear gradient (0 to 100%) of ion exchange buffer B (Tris–HCl 10 mM, pH 7.4, 1 M NaCl, 5 mM EDTA, 1 mM benzamidine, 1 mM PMSF, 1 μ g/mL pepstatin A and 5 mM β -mercaptoethanol). Fractions enriched with the full-length protein were pooled, concentrated to 2 mL using a centrifugal filter device and loaded onto a Superdex 75 16/60 gel-filtration column. Isocratic elution was performed with gel-filtration buffer (10 mM Tris–HCl, pH 7.4, 150 mM NaCl, 5 mM EDTA, 1 mM benzamidine, 1 mM PMSF, 1 μ g/mL pepstatin A and 5 mM β -mercaptoethanol) at a flow rate of 1 mL/min. Gel-filtration of samples prepared for SAXS analyses was performed with a Superdex 75 10/30 column. Tap42 was purified using the same strategy except that imidazole was not added to the affinity buffer. Ion exchange and gel-filtration chromatography were performed using an ÄKTA-FPLC system (Amersham Biosciences) at 25 °C. Purification efficiency was analyzed by SDS-PAGE [28] using a Bio-Rad Mini-Protein III system.

2.4. Purification of $\alpha 4\Delta 222$

The truncated protein $\alpha 4\Delta 222$ was expressed without any affinity tag and was purified by conventional chromatographic techniques using an ÄKTA-FPLC system (Amersham Biosciences). Bacterial cells from 2-L LB cultures were suspended in 20 mL DEAE buffer A (10 mM Tris–HCl, pH 7.4, 10 mM NaCl, 1 mM EDTA, 0.5 mM PMSF, 5% glycerol). Cell extracts were prepared as described above and loaded at a flow rate of 2 mL/min onto a 45-mL DEAE-Sepharose column pre-equilibrated with DEAE buffer A. The column was washed with 90 mL DEAE buffer A and eluted with a 450-mL linear gradient of 0–30% DEAE buffer B (10 mM Tris–HCl, 1 M NaCl, 1 mM EDTA, 0.5 mM PMSF, pH 7.4) at a flow rate of 2.5 mL/min. Fractions enriched with $\alpha 4\Delta 222$ were pooled and brought to 1.2 M $(\text{NH}_4)_2\text{SO}_4$ buffered with 20 mM Tris–HCl pH 7.4. This sample was loaded at a flow rate of 2 mL/min onto a 14 mL Phenyl-Sepharose HP column pre-equilibrated with Phenyl buffer A (20 mM Tris–HCl, pH 7.4, 1.2 M $(\text{NH}_4)_2\text{SO}_4$, 0.5 mM EDTA, 5 mM β -mercaptoethanol). The column was washed with 28 mL of Phenyl buffer A and eluted with a 210 mL linear gradient of 0–60% Phenyl buffer B (20 mM Tris–HCl, pH 7.4, 0.5 mM EDTA, 5 mM β -mercaptoethanol). $\alpha 4\Delta 222$ enriched fractions were pooled and dialyzed against Q-Sepharose buffer A (10 mM Tris–HCl, pH 8.5, 10 mM NaCl, 0.5 mM EDTA, 5 mM β -mercaptoethanol) and loaded at a flow rate of 2 mL/min onto a 16 mL Q-Sepharose HP column pre-equilibrated with the same buffer. The column was washed with 32 mL Q-Sepharose buffer A and $\alpha 4\Delta 222$ was eluted with an 80 mL gradient of 0–100% Q-Sepharose buffer B (10 mM Tris–HCl, pH 8.5, 1 M NaCl, 0.5 mM EDTA, 5 mM β -mercaptoethanol). Samples used for SAXS measurements were dialyzed against the gel-filtration buffer described above and concentrated using a centrifugal filter device.

2.5. Protein quantitation

Protein concentration was estimated from direct absorbance at 280 nm in guanidine buffer (6.0 M guanidine–HCl, 20 mM sodium phosphate buffer, pH 6.55) considering extinction molar coefficients of $\epsilon_{280} = 36840 \text{ M}^{-1} \text{ cm}^{-1}$ for $\alpha 4$, $\epsilon_{280} = 14650 \text{ M}^{-1} \text{ cm}^{-1}$ for $\alpha 4\Delta 236$ and $\alpha 4\Delta 222$, and $\epsilon_{280} = 26740 \text{ M}^{-1} \text{ cm}^{-1}$ for Tap42. For Tap42tr and $\alpha 4$ tr, absorbance was measured in 10 mM sodium phosphate buffer pH 7.2 and 10 mM NaCl, and concentration was calculated considering $\epsilon_{280} = 13410$ and $15930 \text{ M}^{-1} \text{ cm}^{-1}$, respectively.

2.6. Circular dichroism (CD) and sequence analysis

Far UV CD spectra were recorded on a Jasco-810 spectropolarimeter using a Peltier system PFD 425S for temperature control. 3 to 10 μ M protein samples were prepared in 10 mM sodium phosphate buffer pH 7.2 and 10 mM NaCl. CD spectra were acquired at 20 °C using a 1-mm path length cell at 0.5-nm intervals over the wave length range from 190 to 260 nm. Ellipticity is reported as the mean residual ellipticity $[\theta]$ (deg $\text{cm}^2 \text{ dmol}^{-1}$). Samples were subjected to thermal unfolding from 20 °C to 80 °C with spectra collected at 5 °C intervals. The loss of secondary structure was followed by measuring the ellipticity at 222 nm using 0.2 °C intervals. Midpoint transition temperatures were calculated

as the center of the Gaussian fit of the first derivative of the denaturation curves. Refolding assays were started at 80 °C and the temperature lowered to 20 °C with concomitant acquisition of the ellipticity at 222 nm using 0.2 °C intervals. The α -helix content was estimated using the relation described by Gans et al [29]:

$$[\theta]_{222} = [(n - 4.6)(-40,000)]/n \quad (1)$$

where n is the number of amino acid residues of the protein. Prediction of secondary structure elements was performed using PHD [30] (www.embl-heidelberg.de/predictprotein/predictprotein.html). Unstructured regions were estimated using PONDR (www.pondr.com, with VL-XT predictor) [31]. Mean net charge versus mean hydrophobicity was plotted as described by Uversky et al. [32].

2.7. Limited proteolysis

0.3 μ L of trypsin (2 mg/mL in 50 mM NH_4HCO_3 pH 8.0) were added to 120 μ L of protein sample at a concentration of 0.5 mg/mL. The buffers used in limited proteolysis reactions were phosphate buffer pH 7.2 for $\alpha 4$ and its N-terminal truncated domain ($\alpha 4\text{tr}$) and Tris buffer pH 7.0 (20 mM Tris-HCl, 10 mM NaCl) for Tap42 and its N-terminal domain (Tap42tr). The reactions were kept at room temperature and 20 μ L aliquots were taken at various times, mixed with an equal volume of 2 $\times$ SDS-PAGE sample buffer containing 1 mM PMSF and heated to 95 °C for 5 min to inactivate the enzyme. The samples were analyzed by SDS-PAGE.

2.8. SAXS measurements and data analysis

SAXS data were collected at the SAS beamline of the Brazilian Synchrotron Light Laboratory (LNLS) with a 1.488 Å wavelength and a capillary sample holder of 1.5 mm diameter [33]. Protein samples and the molecular weight standard (BSA, 5 mg/mL) were prepared in gel-filtration buffer as described above. Sample to detector distances were 750.2 mm (long) and 446.6 mm (short). Tap42, $\alpha 4$ and $\alpha 4\Delta 236$ contain an N-terminal histidine tag. The following sample concentrations were analyzed: 5 mg/mL BSA, 2 mg/mL $\alpha 4$, 6.9 mg/mL Tap42, 1.7 mg/mL $\alpha 4\Delta 236$, 2 mg/mL $\alpha 4\Delta 222$ and 5 mg/mL $\alpha 4\Delta 222$ (long distance), and 1 mg/mL $\alpha 4$, 21 mg/mL $\alpha 4\Delta 236$ and 26 mg/mL $\alpha 4\Delta 222$ (short distance). Subtraction of parasitic scattering, calculation of the scattering vector q ($q = (4\pi/\lambda)\sin\theta$, 2θ being the scattering angle and λ the wavelength of the incident monochromatic beam) and error calculations were performed using the program TRATID [34]. Fitting of the experimental data and calculation of the pair-distance distribution functions were performed using the GNOM program package [35]. Model calculations were performed using the programs DAMMIN [36] and GASBOR [37]. The program DAMMIN uses simulated annealing optimization to generate a bead atom model that gives the best fit of the scattering intensity. The resulting dummy atom model represents the external shape of the scattering particle. To increase the reliability of the results, the final models for the dummy atom modeling were obtained by a spatial average of ten independent models, calculated with the program DAMAVER [38]. The program GASBOR [37] uses a chain of dummy residues, with the same length of the protein sequence, to represent the protein backbone. Following the same optimization procedure, the program generates a dummy chain model that gives the general shape of the scattering particle and can also give information about the general folding of the studied particle.

3. Results

3.1. Expression and purification of Tap42, $\alpha 4$ and of its C-terminal truncated forms

Expression of his-tagged $\alpha 4$ in *E. coli* BL21 (DE3) yielded approximately 12 mg of soluble protein per liter of culture. His-tagged Tap42 was expressed in *E. coli* M15 and yielded approximately 50 mg of soluble protein per liter of culture. $\alpha 4$ and Tap42 undergo proteolysis after bacterial lysis, generating

fragments of approximately 30 kDa (Fig. 2A, C), which co-purify with each full length protein in metal-chelating affinity columns. The proteolysis fragments correspond to the N-terminal region and co-purify by binding to the affinity column through the N-terminal histidine tag. Thus, the cleavage should

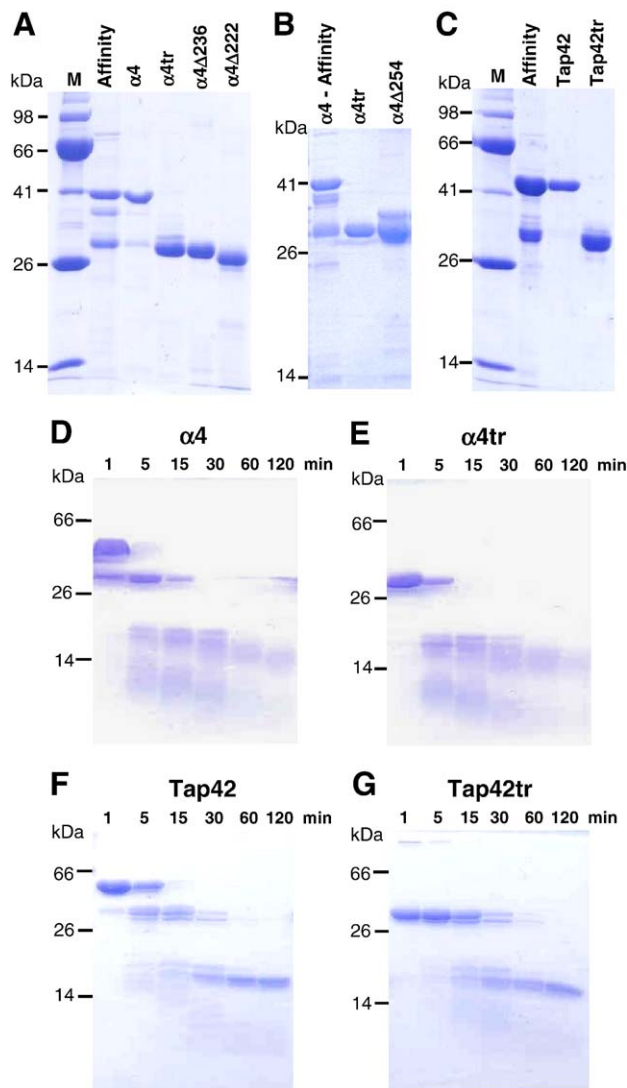


Fig. 2. Coomassie-stained SDS-PAGE gels showing purified proteins and limited proteolysis products. (A) Purification of full-length $\alpha 4$, $\alpha 4\text{tr}$, $\alpha 4\Delta 236$ and $\alpha 4\Delta 222$. The proteins are indicated above each lane. Affinity refers to $\alpha 4$ and $\alpha 4\text{tr}$ purified by ion exchange and gel-filtration chromatography. Lanes labeled $\alpha 4\Delta 236$ and $\alpha 4\Delta 222$ indicate the respective purified protein. (B) Purification of $\alpha 4\Delta 254$. Samples of the affinity purified $\alpha 4$ and of the purified $\alpha 4\text{tr}$, indicated by $\alpha 4$ -Affinity and $\alpha 4\text{tr}$, respectively, were also loaded on the gel for comparison. (C) Purification of Tap42 and Tap42tr. Affinity indicates the sample obtained after affinity chromatography and Tap42 and Tap42tr indicate the full-length and truncated proteins after fractionation by ion-exchange chromatography. The size of the molecular weight markers (lane M) are indicated on the left side of each panel. (D–G) Limited proteolysis analysis of $\alpha 4$ (D); $\alpha 4\text{tr}$ (E), Tap42 (F) and Tap42tr (G). The molecular weights of the limited proteolysis products of $\alpha 4$, $\alpha 4\text{tr}$, Tap42 and Tap42tr treated with trypsin are in agreement with those obtained during the purification process. The reaction was prepared with a 1:100 (w:w) protein:trypsin ratio and aliquots were removed at various times, indicated above each lane. Position of the molecular weight markers are indicated on the left side of each panel.

occur at corresponding sites in both proteins. His-tagged $\alpha 4$ and Tap42 were purified using the same strategy, consisting of an immobilized metal affinity chromatography, followed by anion exchange and size-exclusion chromatography and relying on protease inhibitors to minimize degradation. After the affinity chromatography, the full-length and the proteolysis products of both proteins are present in similar amounts, being separated by anion exchange chromatography under the same buffer conditions (Fig. 2A, C). Interestingly, the elution profiles of both proteins at this step are very similar (not shown), suggesting that full-length $\alpha 4$ and Tap42 as well as their respective proteolysis products have similar charge distribution. The purified $\alpha 4$ and Tap42 proteolysis products will be referred to as $\alpha 4$ tr and Tap42tr, respectively.

Three deletion mutants of $\alpha 4$ were constructed to tentatively map the site of proteolysis and to generate truncated proteins for structural studies. The apparent size of the proteolysis product on SDS-PAGE is ~ 30 kDa. Taking into account this information and secondary structure prediction, indicating a large coil region between amino acids 221 and 267 (Fig. 1), which could be a possible target for proteases, three truncated proteins were designed and cloned: $\alpha 4\Delta 222$, containing amino acids 1 to 222; $\alpha 4\Delta 236$, comprising amino acids 1 to 236 and $\alpha 4\Delta 254$, containing amino acids 1 to 254 (Fig. 1). Interestingly, $\alpha 4\Delta 254$ undergoes proteolysis, resulting in an N-terminal fragment which has the same apparent molecular weight of the one resulting from proteolysis of the full-length protein (Fig. 2B). $\alpha 4\Delta 236$ and $\alpha 4\Delta 222$ do not undergo proteolysis, showing that the cleavage site is located between residues 236 and 254. Expression of $\alpha 4\Delta 236$ and $\alpha 4\Delta 222$ yields significantly more soluble protein than full-length $\alpha 4$. Due to its stability and high level of expression, $\alpha 4\Delta 222$ could be easily purified by conventional chromatographic techniques, avoiding the need for a hexahistidine tag. It was first captured from the *E. coli* cell extracts by anion exchange on DEAE-Sepharose and further purified by hydrophobic interaction chromatography (Phenyl-Sepharose), after which $\alpha 4\Delta 222$ was more than 90% pure. The final step consisted of another anion exchange chromatography using a stronger anion exchanger (Q-Sepharose) and a buffer with higher pH to remove the remaining contaminants (Fig. 2A).

3.2. Limited proteolysis

Tap42 and $\alpha 4$ were subjected to limited digestion with trypsin to verify the presence of proteolytically unstable regions (Fig. 2D–G). Their respective N-terminal regions generated by proteolysis during the purification process were digested in parallel as controls (Fig. 2E, G). As expected, the full-length proteins are first converted to a fragment of ~ 30 kDa (Fig. 2D, F), indicating that the C-terminus is quickly degraded. Following conversion of the full length proteins to the ~ 30 kDa fragments, the kinetics of degradation as well as the pattern of low molecular weight fragments generated is very similar for both $\alpha 4$ and $\alpha 4$ tr, and for Tap42 and Tap42tr. Cleavage by trypsin to generate the ~ 30 kDa fragment should occur at the same region cleaved by the *E. coli* proteases found in the lysates, indicating that proteolysis is not specific. In

conclusion, this experiment shows that $\alpha 4$ and Tap42 have C-terminal regions highly sensitive to protease attack. The functional implications of this sensitivity, however, remain unclear.

3.3. Secondary structure and thermally-induced unfolding analyses using circular dichroism

The CD spectra of $\alpha 4$, Tap42 and their respective proteolysis products are very similar, showing that both proteins contain equivalent secondary structure elements, further supporting the structural similarity of these two proteins (Fig. 3A, B). Although the spectra of the full-length proteins and of the proteolysis products are typical of α -helical proteins, the $\alpha 4\Delta 236$ and $\alpha 4\Delta 222$ spectra showed a relatively higher α -helical content (Fig. 3C), indicating that the C-terminal region is less structured, consistently with the secondary structure prediction (Fig. 1). An estimation using the equation described by Gans et al. [29] suggests that the α -helix content of $\alpha 4\Delta 236$ and $\alpha 4\Delta 222$ is 50% and 41%, respectively, while $\alpha 4$ contains 31% of α -helices, which is consistent with the presence of a coil-rich region in the C-terminus of $\alpha 4$.

Thermal-induced unfolding followed by CD was performed to determine whether the higher proteolysis resistance of the N-terminal region would imply a higher stability for the truncated forms as compared to the respective full-length proteins. Unfolding of Tap42 revealed a midpoint transition temperature of 48 °C (Fig. 4A). Tap42tr aggregates following temperature increase above 45 °C, making it impossible to determine its transition temperature. Thermal unfolding leads Tap42 to acquire a β -sheet-like structure while Tap42tr shows a total loss CD signal due to aggregation (data not shown). The thermal effect observed for $\alpha 4$ and for its C-terminal truncated products also indicates a change to a β -sheet-like folding (data not shown). The unfolding kinetics of these proteins shows a gradual loss of secondary structure, starting from 10 °C, with $\alpha 4$ tr showing a slightly lower midpoint transition temperature (50 °C), whereas $\alpha 4$ (54 °C), $\alpha 4\Delta 236$ (55.3 °C) and $\alpha 4\Delta 222$ (55.2 °C) showed similar transition temperatures (Fig. 4B). This finding was intriguing because the C-terminal region was expected to have an effect on the unfolding of $\alpha 4$, unless it was already unfolded prior to temperature increase. Unfolding was not reversible for both $\alpha 4$ and Tap42, as well as for truncated proteins (data not shown).

3.4. SAXS results and structural models

Small angle scattering curves for Tap42, $\alpha 4$ (his-tagged $\alpha 4$) and its deletion mutants $\alpha 4\Delta 222$ and $\alpha 4\Delta 236$ (his-tagged $\alpha 4\Delta 236$) as well as their corresponding pair distance distribution functions $p(r)$ are shown in Fig. 5. Measurements at short and long distances for several concentrations were combined to produce reliable SAXS patterns within the scattering range recorded in the experiments. The $p(r)$ functions indicate that Tap42 and $\alpha 4$ have a very elongated shape, while $\alpha 4\Delta 236$ and $\alpha 4\Delta 222$ are slightly elongated. The radii of gyration were calculated from the second moment of the $p(r)$

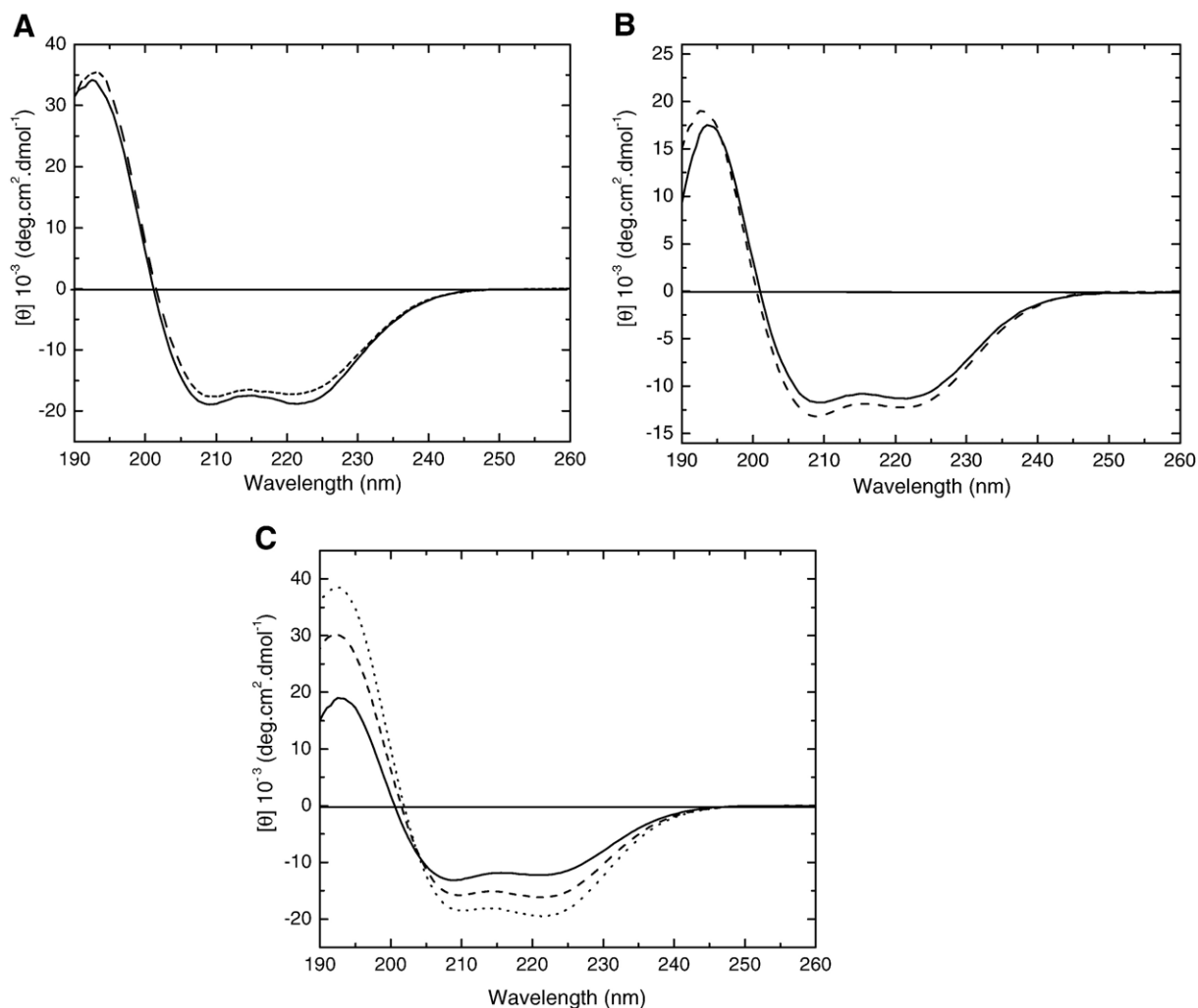


Fig. 3. Circular dichroism analyses. CD spectra were acquired at 20 °C with protein samples at 3 to 10 μ M in 10 mM sodium phosphate buffer pH 7.2 and 10 mM NaCl as described under Materials and methods. (A) CD spectra of α 4tr (solid line) and Tap42tr (dashed line). (B) CD spectra of α 4 (solid line) and Tap42 (dashed line). (C) CD spectra of α 4 (solid line), α 4 Δ 236 (dashed line) and α 4 Δ 222 (dotted line).

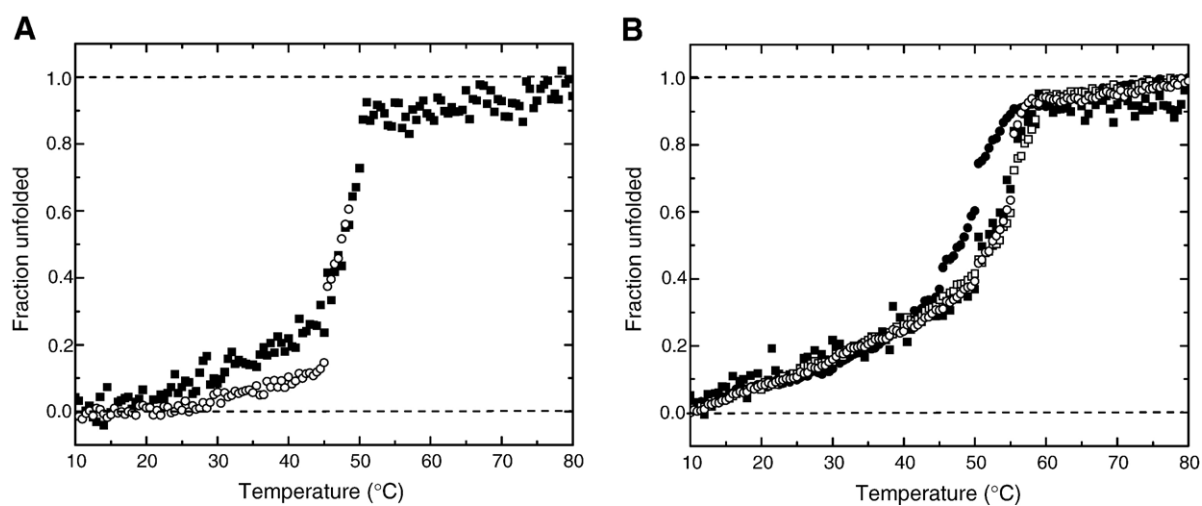


Fig. 4. Analysis of thermal-induced unfolding of α 4, Tap42 and truncated proteins. Unfolding was followed by measuring the ellipticity at 222 nm using 0.2 °C intervals in the temperature range from 20 °C to 80 °C. Proteins (5 μ M) were prepared in 10 mM sodium phosphate buffer pH 7.2 and 10 mM NaCl. (A) Thermal unfolding of Tap42 (■) and Tap42tr (○). Tap42tr data are shown only up to ~50 °C, since it aggregates above this temperature. (B) Thermal unfolding of α 4 (■), α 4tr (●), α 4 Δ 236 (□) and α 4 Δ 222 (○).

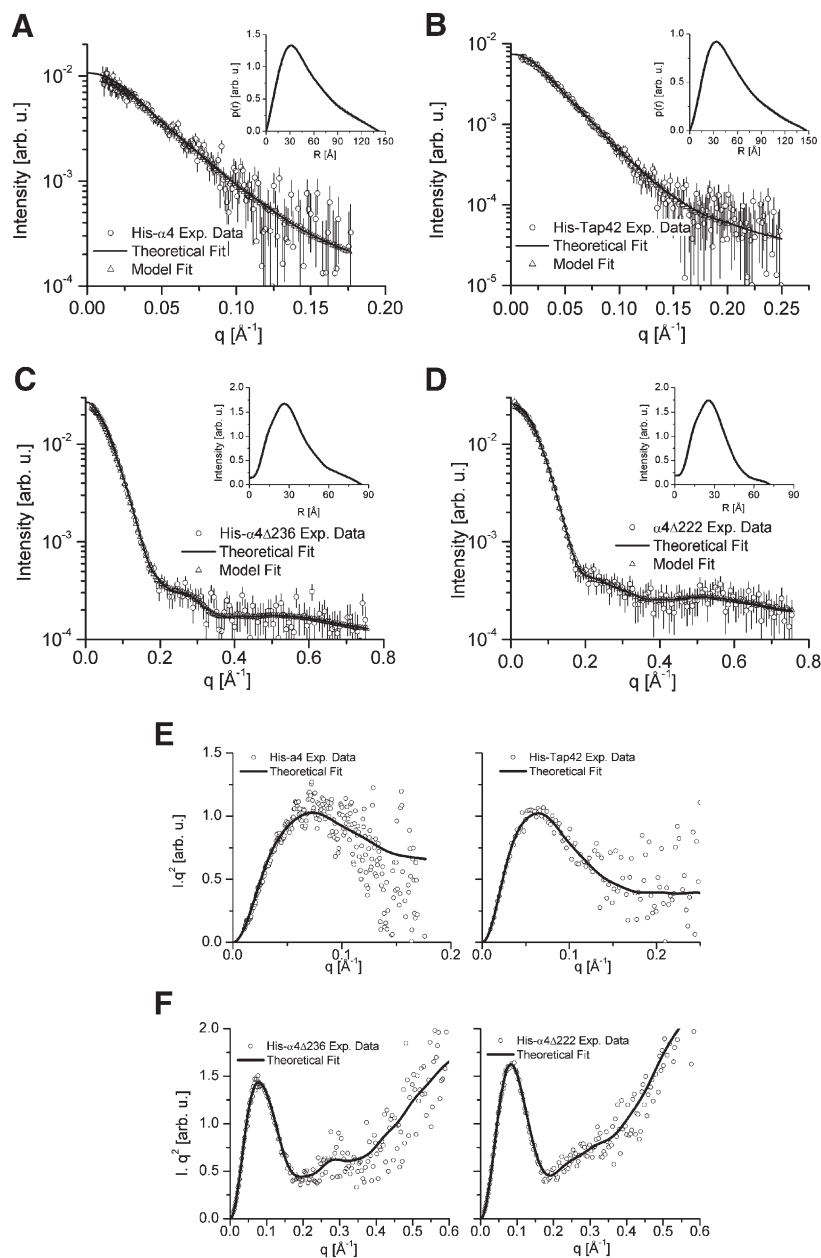


Fig. 5. Experimental solution scattering curves, theoretical (GNOM) and model curve-fittings for $\alpha 4$ (A), Tap42 (B), $\alpha 4\Delta 236$ (C) and $\alpha 4\Delta 222$ (D). Circles and error bars indicate the experimental data. Solid lines indicate the GNOM theoretical fit. Open triangles indicate ab initio model fits. The inset in each graph shows the pair-distance distribution function. Kratky plots for the four proteins are shown in panels E and F.

function. The values obtained were: 42.8 ± 0.7 $\text{\AA}$ for Tap42, 41.2 ± 0.8 $\text{\AA}$ for $\alpha 4$, 25.7 ± 0.2 $\text{\AA}$ for $\alpha 4\Delta 236$ and 21.6 ± 0.3 $\text{\AA}$ for $\alpha 4\Delta 222$. The maximum dimensions for the proteins were determined from the value where the $p(r)$ function reaches zero, these values being ~ 147 $\text{\AA}$ (Tap42), ~ 142 $\text{\AA}$ ($\alpha 4$), ~ 85 $\text{\AA}$ ($\alpha 4\Delta 236$) and ~ 72 $\text{\AA}$ ($\alpha 4\Delta 222$).

Molecular weight estimations could be obtained from the extrapolation of the scattering intensities to zero angle performed by the theoretical fit with program GNOM. Using bovine serum albumin (BSA, 66 kDa) as a secondary standard, we obtained molecular weight values that were in good agreement with those expected for the three proteins according to their primary sequences. This showed that the

three proteins were monomeric in solution, assuring the monodispersity of the samples. The values obtained were 43 ± 1 kDa for Tap42, 45 ± 5 kDa for $\alpha 4$, 28 ± 2 kDa for $\alpha 4\Delta 236$ and 23 ± 1 kDa for $\alpha 4\Delta 222$.

The so called Kratky plots ($I \cdot q^2$ vs. q) are a very useful tool to describe the degree of compactness of scattering particles [39]. For all studied proteins, the behavior of the Kratky plots suggests the presence of a certain degree of non compactness which can indicate the presence of flexible domains. From the comparison of the Kratky plots up to $q = 0.2$ $\text{\AA}^{-1}$, it can be clearly seen that the graphs for the deletion mutants have a higher and sharper maximum than the full-length proteins (Fig. 5E, F). Based on CD results, this can be interpreted as an

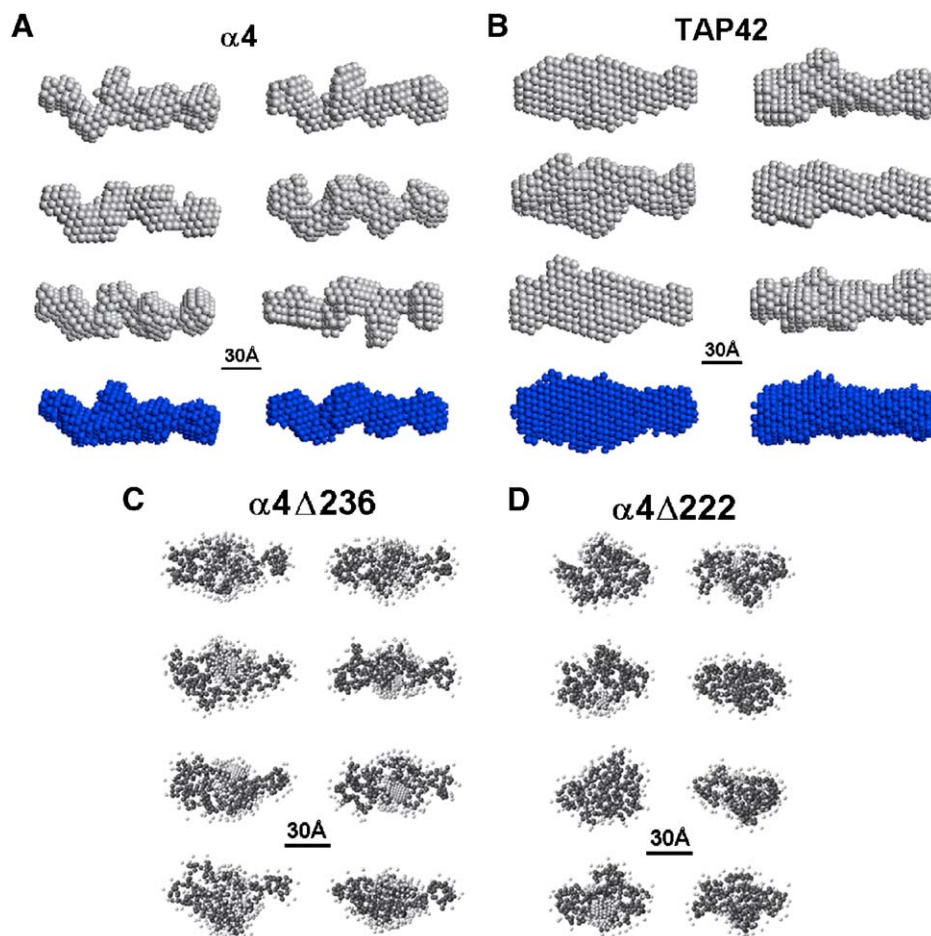


Fig. 6. Representation of the structural models constructed by ab initio modeling. Dummy atom models for $\alpha 4$ (A) and Tap42 (B) were calculated using the program DAMMIN. Three independent calculations are shown in two lengthwise views. The view on the right is rotated 90° relative to the view on the left (gray shade). The fourth model (blue shade) is an average of all independent models calculated. Dummy chain models for $\alpha 4\Delta 236$ and $\alpha 4\Delta 222$ were calculated using the program GASBOR. Four independent higher resolution models are presented in panels C and D, respectively. Models on the right are rotated 90° relative to the ones on the left. An average structure is not presented in this case to avoid loss of resolution.

indication of higher coil content in the C-terminal part of $\alpha 4$ and probably also in Tap42.

It is interesting to note that $\alpha 4\Delta 236$ and $\alpha 4\Delta 222$ could be obtained at the concentrations of 21 and 26 mg/mL, respectively, which resulted in a good scattering signal even for high q values. The maximum value of q obtained for these proteins was $q_{\max} = 0.76 \text{ \AA}^{-1}$, which means a resolution ($R = 2\pi/q_{\max}$) of $R = 8.4 \text{ \AA}$. As known from SAXS theory and modeling [40], higher q data increase the information content of the scattering curves, decreasing the ambiguity of the calculated models. The parameter N_S (number of Shannon channels)² is an indicator of the resolution that has been reached in each case. For $\alpha 4\Delta 236$ and $\alpha 4\Delta 222$ we obtained $N_S = 20$ and 17, respectively. This allowed the construction of structural models with a higher resolution for the N-terminus of the protein. For the full-length proteins, Tap42 and $\alpha 4$, it was not possible to reach equally high q values due to the low concentration of the

samples (Fig. 5). In this case, the number of Shannon channels obtained were $N_S = 9$ and 8, respectively.

The next step in this investigation was the calculation of low resolution structural models. The choice of the correct modeling approach depends on the amount of information available in each case. For the full length proteins, $\alpha 4$ and Tap42, a dummy atom modeling procedure was applied [36]. This approach (recommended for $q < 0.2 \text{ \AA}^{-1}$) gives the so called “shape scattering” and the resulting model represents the general shape of the scattering particle [41]. For the deletion mutants $\alpha 4\Delta 236$ and $\alpha 4\Delta 222$, we applied a dummy chain modeling approach [37]. This approach (recommended for $q > 0.25 \text{ \AA}^{-1}$) can give indication about quaternary structure, domain arrangement and also general folding of the studied proteins [41, 42]. In both cases, the resulting model does not correspond to a unique solution [40]. For the dummy atom models, an average of the models can give the most probable structure [38]. Averaged models were not calculated for the dummy chain models of $\alpha 4\Delta 236$ and $\alpha 4\Delta 222$ to get a better insight of the general folding and four resulting models are shown in Fig. 6C and D, respectively. The best solutions in each case are shown in this

² $N_S = (q_{\max} - q_{\min})D_{\max}/\pi$, $q_{\max}$ ($q_{\min}$) being the maximum (minimum) experimental q values and $D_{\max}$ the maximum dimension of the particle.

figure. For $\alpha 4$ and Tap42, the most probable structure is also shown (Fig. 6A, B). From the comparison of the resulting models, some conclusions can be drawn. $\alpha 4$ and Tap42 have very elongated shapes, being the overall model shape similar for both proteins. Due to the low resolution of the models, it is not possible to identify the position of the N- or C-terminus in the models. For the deletion mutants representing the N-terminus, the corresponding models are clearly less ambiguous, which is a result of the higher amount of information obtained from the experimental data. Interestingly, an extension can be observed for the $\alpha 4\Delta 236$ models (Fig. 6C) that is not observed for the $\alpha 4\Delta 222$ models (Fig. 6D). This difference can be explained by the fact that $\alpha 4\Delta 236$ contains 20 additional amino acids at the N-terminus and 14 at the C-terminus relative to $\alpha 4\Delta 222$, representing an additional mass of 3671 Da, equivalent to 14% of $\alpha 4\Delta 222$ molecular mass.

4. Discussion

$\alpha 4$ and Tap42 are conserved proteins sharing functional activities involving regulation of the PP2A-related phosphatases in the context of the rapamycin-sensitive pathway. Recombinant $\alpha 4$ and Tap42 show similar chromatographic behavior and are highly susceptible to proteolysis that degrades the C-terminal region, resulting in ~ 30 kDa stable N-terminal domains. This high sensitivity to proteolysis imposes the use of high amounts of protease inhibitors and additional chromatographic steps to obtain purified proteins. Murine recombinant $\alpha 4$ also undergoes proteolysis on its C-terminus [16] showing that this instability is typical of proteins of this family. High susceptibility of the $\alpha 4$ and Tap42 C-termini to proteolysis was demonstrated by limited trypsin digestion (Fig. 2D–G). In this experiment, both proteins are quickly converted to a ~ 30 kDa product similar to the one generated by the *E. coli* proteases found in the extracts. The proteolysis products contain intact N-termini as they bind to the affinity column through the N-terminal histidine tag and co-purify with the full-length proteins. The C-terminus of the $\alpha 4$ proteolysis product generated by *E. coli* proteases was mapped to the region between amino acids 236 and 254, indicating that part of the predicted coil region remains as part of the proteolysis resistant product. CD studies confirmed that $\alpha 4$ and Tap42 show high α -helical content and the nearly overlapping spectra further indicate their structural similarity. Surprisingly, the high resistance to proteolysis and the high solubility shown by $\alpha 4\Delta 236$ and $\alpha 4\Delta 222$ did not correspond to higher thermal stability, indicating that the C-terminal region does not affect the unfolding mechanism of $\alpha 4$, possibly because it is naturally unstructured.

The N-terminal regions of $\alpha 4$ are more compact and well-structured than the full-length $\alpha 4$, although the SAXS data indicate that a certain degree of flexibility is also shown by the truncated proteins. Based on these data, and from the comparison between $\alpha 4$ and Tap42, it is reasonable to suppose that the structural model with prolate shape obtained in both cases is representative of all proteins of this family. The results obtained by limited proteolysis and SAXS analyses support the

hypothesis that the C-terminal regions of $\alpha 4$ and Tap42 are flexible and could be intrinsically unstructured. Natively unfolded proteins or protein segments have been described for a large number of proteins, including transcriptional and translational regulators, factors involved in protein–protein and interactions and factors showing nucleic acid binding activity [27,43–45]. Secondary structure prediction indicates that the C-terminal regions of $\alpha 4$ and Tap42 are formed mainly by coil and segments without reliable prediction. Prediction of intrinsically disordered regions [31] does find high scores for both C-terminal regions but they are not clearly distinct from the scores observed for other regions of $\alpha 4$ and Tap42 (data not shown). Uversky et al. [32] have used the relation between the normalized net charge versus mean hydrophobicity to show that naturally folded and intrinsically unstructured proteins fall in different regions of charge-hydrophobicity phase space, revealing that intrinsically unstructured proteins show a combination of high mean net charge and low hydrophobicity. Interestingly, by plotting the mean net charge versus mean hydrophobicity, we observed that the N-terminal regions corresponding to the protease resistant domains are localized in the phase space of natively folded proteins, whereas full-length $\alpha 4$ and Tap42 and especially the C-terminal regions are localized in the phase space of intrinsically unstructured proteins (Fig. 7).

The presence of natively unstructured domains, which may fold and stabilize upon substrate binding, is a common theme in regulatory and signaling proteins and may allow them to recognize different targets with high specificity [27,32,43–45]. Therefore, the structural flexibility of $\alpha 4$ and Tap42 may be important for them to exert their cellular functions by interacting with other proteins and regulating their activity. The $\alpha 4$ /Tap42 best described interacting partner is the catalytic subunit of PP2A, which binds $\alpha 4$ on its N-terminal domain, mapped to residues 94–202 on murine $\alpha 4$ [18]. Association with $\alpha 4$ augments phosphatase activity towards phosphorylase *a* [16], myelin basic protein [16,18], histone [18], and translation elongation factor 2 [17], but diminishes its activity toward *p*-nitrophenyl phosphate [46] and eIF-4E-binding

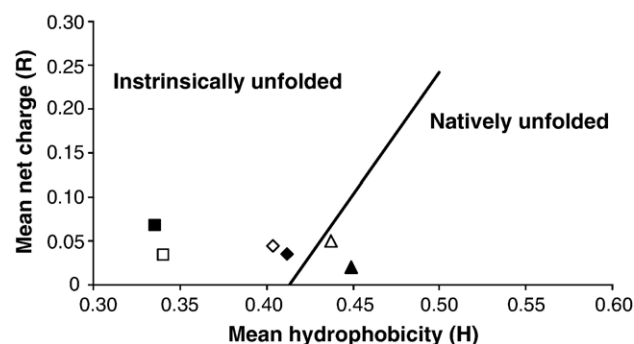


Fig. 7. Analysis of the net charge and hydrophobicity of $\alpha 4$ and Tap42. The mean net charge (R) and mean hydrophobicity were plotted for full-length $\alpha 4$ (◇) Tap42 (◆), the N-terminal regions of $\alpha 4$ (△, residues 1–222) and of Tap42 (▲, residues 1–247) and for the C-terminal regions of $\alpha 4$ (□, residues 223–339) and of Tap42 (■, residues 248–366). The continuous black line ($R=2.785H-1.151$) represents the boundary between natively unfolded and folded proteins determined by Uversky et al. [40].

protein [21]. In spite of the discrepancy among these published data, it is clear that the regulation of PP2A by $\alpha 4$ is not a simple question of activation or inhibition, but rather of substrate specificity. The data obtained in this work suggest a model in which $\alpha 4$ binds the phosphatase catalytic subunit by its rigid N-terminal domain, while its flexible C-terminus interacts with diverse substrates and thus mediates their interaction with the phosphatase. This model could explain the mechanism of regulation of MID1 dephosphorylation, which is catalyzed by $\alpha 4$ -bound PP2Ac, in which the phosphatase is targeted to its substrate by the interaction between the C-terminus of $\alpha 4$ and the B-box domain of MID1 [12,13]. Another feature conferred by the structural flexibility of some proteins is their easy degradation, which might be essential for regulatory proteins whose cellular levels must change rapidly in response to environmental stimuli. Since Tap42 and $\alpha 4$ were shown to be degraded preferentially from their C-termini in vitro, it would be interesting to investigate the consequences of this degradation kinetics in vivo.

Acknowledgements

We thank Tereza C. Lima Silva, Adriana C. Alves, Luciana R. Camillo, Romênia R. Domingues, Leandro C. Vieira, Helder A. Silva and Zildene G. Correa for technical assistance and Yraima Cordeiro for helping with Tap42 CD experiments. J.H.C.S., C.L.P.O., T.A.P. and F.R.G.C. were recipients of FAPESP fellowships. C.L.P.O. was also a recipient of CNPq fellowship. This work was supported by FAPESP grant 00/02788-4 (N.I.T.Z.), the FAPESP SMolBNet and CEPID/CBME programs, and by the National Structural Biology Program from the LNLS/CNPq/MCT.

References

- [1] T. Schmelzle, M.N. Hall, TOR, a central controller of cell growth, *Cell* 103 (2000) 253–262.
- [2] N. Hay, N. Sonenberg, Upstream and downstream of mTOR, *Genes Dev.* 18 (2004) 1926–1945.
- [3] C.J. Di Como, K.T. Arndt, Nutrients, via the TOR proteins, stimulate the association of Tap42p with type 2A phosphatases, *Genes Dev.* 10 (1996) 1904–1916.
- [4] K. Kuwahara, T. Matsuo, J. Nomura, H. Igarashi, M. Kimoto, S. Inui, N. Sakaguchi, Identification of a 52-kDa molecule (p52) coprecipitated with the Ig receptor-related MB-1 protein that is inducibly phosphorylated by the stimulation with phorbol myristate acetate, *J. Immunol.* 152 (1994) 2742–2752.
- [5] C.K. Too, Differential expression of elongation factor-2, $\alpha 4$ phosphoprotein and Cdc5-like protein in prolactin-dependent/independent rat lymphoid cells, *Mol. Cell. Endocrinol.* 131 (1997) 221–232.
- [6] L.T. Binh, K. Oono, Molecular cloning and characterization of genes related to chilling tolerance in rice, *Plant Physiol.* 99 (1992) 1146–1150.
- [7] D.M. Harris, T.L. Myrick, S.J. Rundle, The *Arabidopsis* ortholog of yeast TAP42p and mammalian $\alpha 4$ binds to the catalytic subunit of protein phosphatase 2A and is induced by chilling, *Plant Physiol.* 121 (1999) 609–617.
- [8] S. Inui, K. Kuwahara, J. Mizutani, K. Maeda, T. Kawai, H. Nakayasu, N. Sakaguchi, Molecular cloning of a cDNA clone encoding a phosphoprotein component related to the Ig receptor-mediated signal transduction, *J. Immunol.* 154 (1995) 2714–2723.
- [9] S. Inui, K. Maeda, D.R. Hua, T. Yamashita, H. Yamamoto, E. Miyamoto, S. Aizawa, N. Sakaguchi, BCR signal through $\alpha 4$ is involved in S6 kinase activation and required for B cell maturation including isotype switching and V region somatic hypermutation, *Int. Immunol.* 14 (2002) 177–187.
- [10] D.R. Hua, S. Inui, T. Yamashita, K. Maeda, K. Takagi, J. Takeda, N. Sakaguchi, T cell-specific gene targeting reveals that $\alpha 4$ is required for early T cell development, *Eur. J. Immunol.* 33 (2003) 1899–1906.
- [11] M. Kong, C.J. Fox, J. Mu, L. Solt, A. Xu, R.M. Cinalli, M.J. Birnbaum, T. Lindsten, C.B. Thompson, The PP2A-associated protein $\alpha 4$ is an essential inhibitor of apoptosis, *Science* 306 (2004) 695–698.
- [12] J. Liu, T.D. Prickett, E. Elliott, G. Meroni, D.L. Brautigan, Phosphorylation and microtubule association of the Opitz syndrome protein mid-1 is regulated by protein phosphatase 2A via binding to the regulatory subunit $\alpha 4$, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 6650–6655.
- [13] K. Short, B. Hopwood, Z. Yi, T.C. Cox, MID1 and MID2 homo- and heterodimerise to tether the rapamycin-sensitive PP2A regulatory subunit, $\alpha 4$, to microtubules: implications for the clinical variability of X-linked Opitz GBBB syndrome and other developmental disorders, *BMC Cell Biol.* 3 (2002) 1.
- [14] Y. Goldberg, Protein phosphatase 2A: who shall regulate the regulator? *Biochem. Pharmacol.* 57 (1999) 321–328.
- [15] T.D. Prickett, D.L. Brautigan, Overlapping binding sites in protein phosphatase 2A for association with regulatory A and $\alpha 4$ (mTap42) subunits, *J. Biol. Chem.* 279 (2004) 38912–38920.
- [16] K. Murata, J. Wu, D.L. Brautigan, B cell receptor-associated protein $\alpha 4$ displays rapamycin-sensitive binding directly to the catalytic subunit of protein phosphatase 2A, *Proc. Natl. Acad. Sci. U. S. A.* 94 (1997) 10624–10629.
- [17] H. Chung, A.C. Nairn, K. Murata, D.L. Brautigan, Mutation of Tyr307 and Leu309 in the protein phosphatase 2A catalytic subunit favors association with the $\alpha 4$ subunit which promotes dephosphorylation of elongation factor-2, *Biochemistry* 38 (1999) 10371–10376.
- [18] S. Inui, H. Sanjo, K. Maeda, H. Yamamoto, E. Miyamoto, N. Sakaguchi, Ig receptor binding protein 1 ($\alpha 4$) is associated with a rapamycin-sensitive signal transduction in lymphocytes through direct binding to the catalytic subunit of protein phosphatase 2A, *Blood* 92 (1998) 539–546.
- [19] T. Yamashita, S. Inui, K. Maeda, D.R. Hua, K. Takagi, N. Sakaguchi, The heterodimer of $\alpha 4$ and PP2Ac is associated with S6 kinase1 in B cells, *Biochem. Biophys. Res. Commun.* 330 (2) (2005) 439–445.
- [20] J. Chen, R.T. Peterson, S.L. Schreiber, $\alpha 4$ associates with protein phosphatases 2A, 4, and 6, *Biochem. Biophys. Res. Commun.* 247 (1998) 827–832.
- [21] M. Nanahoshi, T. Nishiuma, Y. Tsujishita, K. Hara, S. Inui, N. Sakaguchi, K. Yonezawa, Regulation of protein phosphatase 2A catalytic activity by $\alpha 4$ protein and its yeast homolog Tap42, *Biochem. Biophys. Res. Commun.* 251 (1998) 520–526.
- [22] S. Kloerker, R. Reed, J.L. McConell, D. Chang, K. Tran, R.S. Westphal, B.K. Law, R.J. Colbran, M. Kamoun, K.S. Campbell, B.E. Wadzinski, Parallel purification of three catalytic subunits of the protein serine/threonine phosphatase family (PP2A_c, PP4_c and PP6_c) and analysis of the interaction of PP2A_c with $\alpha 4$ protein, *Protein Expr. Purif.* 31 (2003) 19–33.
- [23] A.L. Fink, Natively unfolded proteins, *Curr. Opin. Struct. Biol.* 15 (2005) 35–41.
- [24] C. Bracken, L.M. Iakoucheva, P.R. Romero, A.K. Dunker, Combining prediction, computation and experiment for the characterization of protein disorder, *Curr. Opin. Struct. Biol.* 14 (2004) 570–576.
- [25] V.N. Uversky, What does it mean to be natively unfolded? *Eur. J. Biochem.* 269 (2002) 2–12.
- [26] J.J. Ward, J.S. Sodhi, L.J. McGuffin, B.F. Buxton, D.T. Jones, Prediction and functional analysis of native disorder in proteins from the three kingdoms of life, *J. Mol. Biol.* 337 (2004) 635–645.
- [27] L.M. Iakoucheva, C.J. Brown, J.D. Lawson, Z. Obradovic, A.K. Dunker, Intrinsic disorder in cell-signaling and cancer-associated proteins, *J. Mol. Biol.* 323 (2002) 573–584.

- [28] F.M. Ausubel, R. Brent, R. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl, *Current Protocols in Molecular Biology*, John Wiley and Sons, New York, 1998.
- [29] P.J. Gans, P.C. Lyu, M.C. Manning, R.W. Woody, N.R. Kallenbach, The helix-coil transition in heterogeneous peptides with specific side-chain interactions: theory and comparison with CD spectral data, *Biopolymers* 31 (1991) 1605–1614.
- [30] B. Rost, PHD: predicting one-dimensional protein structure by profile-based neural networks, *Methods Enzymol.* 266 (1996) 525–539.
- [31] P. Romero, Z. Obradovic, A.K. Dunker, Sequence data analysis for long disordered regions prediction in the calcineurin family, *Genomics Inf.* 8 (1997) 110–124.
- [32] V.N. Uversky, J.R. Gillespie, A.L. Fink, Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* 41 (2000) 415–427.
- [33] L.P. Cavalcanti, I.L. Torriani, T.S. Plivelic, C.L.P. Oliveira, G. Kellermann, R. Neuenschwander, Two new sealed sample cells for small angle X-ray scattering from macromolecules in solution and complex fluids using synchrotron radiation, *Rev. Sci. Instrum.* 75 (2004) 4541–4546.
- [34] C.L.P. Oliveira, TRAT1D- Computer Program for SAXS Data Treatment, LNLS Technical Manual MT01/2003, 2003.
- [35] V. Semenyuk, D.I. Svergun, GNOM—A program package for small-angle scattering data-processing, *J. Appl. Cryst.* 24 (1991) 537–540.
- [36] D.I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, *Biophys. J.* 76 (1999) 2879–2886.
- [37] D.I. Svergun, M.V. Petoukhov, M.H.J. Koch, Determination of domain structure of proteins from X-ray solution scattering, *Biophys. J.* 80 (2001) 2946–2953.
- [38] V.V. Volkov, D.I. Svergun, Uniqueness of ab initio shape determination in small-angle scattering, *J. Appl. Cryst.* 36 (2003) 860–864.
- [39] O. Glatter, O. Kratky, *Small Angle X-ray Scattering*, Academic Press, London, 1982.
- [40] D.I. Svergun, M.H.J. Koch, Small-angle scattering studies of biological macromolecules in solution, *Rep. Prog. Phys.* 66 (2003) 1735–1782.
- [41] K. Bilecen, U.H. Ozturk, A.D. Duru, T. Sutlu, M.V. Petoukhov, D.I. Svergun, M.H. Koch, U.O. Sezerman, I. Cakmak, Z. Sayers, Triticum durum metallothionein. Isolation of the gene and structural characterization of the protein using solution scattering and molecular modeling, *J. Biol. Chem.* 280 (2005) 13701–13711.
- [42] M.V. Petoukhov, D.I. Svergun, Global rigid body modeling of macromolecular complexes against small-angle scattering data, *Biophys. J.* 89 (2005) 1237–1250.
- [43] L.M. Iakoucheva, P. Radivojac, C.J. Brown, T.R. O'Connor, J.G. Sikes, Z. Obradovic, A.K. Dunker, The importance of intrinsic disorder for protein phosphorylation, *Nucleic Acids Res.* 32 (2004) 1037–1049.
- [44] H.J. Dyson, P.E. Wright, Intrinsically unstructured proteins and their functions, *Nat. Rev., Mol. Cell Biol.* 6 (2005) 197–208.
- [45] V.N. Uversky, Natively unfolded proteins: a point where biology waits for physics, *Protein Sci.* 11 (2002) 739–756.
- [46] M. Nanahoshi, Y. Tsujishita, C. Tokunaga, S. Inui, N. Sakaguchi, K. Hara, K. Yonezawa, Alpha4 protein as a common regulator of type 2A-related serine/threonine protein phosphatases, *FEBS Lett.* 446 (1999) 108–112.

8.2- ANEXO II

Scapin, S. M., Carneiro, F. R., Alves, A. C., Medrano, F. J., Guimaraes, B. G., Zanchin, N. I. 2006. The crystal structure of the small GTPase Rab11b reveals critical differences relative to the Rab11a isoform. *J Struct Biol.* 2006 Feb 20; [Epub ahead of print].



The crystal structure of the small GTPase Rab11b reveals critical differences relative to the Rab11a isoform

Sandra M.N. Scapin, Flávia R.G. Carneiro, Adriana C. Alves, F. Javier Medrano,
Beatriz G. Guimarães*, Nilson I.T. Zanchin*

Center for Structural Molecular Biology, Brazilian Synchrotron Light Laboratory, LNLS, P.O. Box 6192, CEP 13084-971, Campinas, SP, Brazil

Received 14 December 2005; received in revised form 13 January 2006; accepted 18 January 2006

Abstract

Rab GTPases constitute the largest family of small monomeric GTPases, including over 60 members in humans. These GTPases share conserved residues related to nucleotide binding and hydrolysis, and main sequence divergences lie in the carboxyl termini. They cycle between inactive (GDP-bound) and active (GTP-bound) forms and the active site regions, termed Switch I and II, undergo the larger conformational changes between the two states. The Rab11 subfamily members, comprising Rab11a, Rab11b, and Rab25, act in recycling of proteins from the endosomes to the plasma membrane, in transport of molecules from the trans-Golgi network to the plasma membrane and in phagocytosis. In this work, we describe Rab11b-GDP and Rab11b-GppNHp crystal structures solved to 1.55 and 1.95 Å resolution, respectively. Although Rab11b shares 90% amino acid identity to Rab11a, its crystal structure shows critical differences relative to previously reported Rab11a structures. Inactive Rab11a formed dimers with unusually ordered Switch regions and missing the magnesium ion at the nucleotide binding site. In this work, inactive Rab11b crystallized as a monomer showing a flexible Switch I and a magnesium ion which is coordinated by four water molecules, the phosphate β of GDP (β -P) and the invariant S25. S20 from the P-loop and S42 from the Switch I are associated to GTP hydrolysis rate. In the active structures, S20 interacts with the γ -P oxygen in Rab11b-GppNHp but does not in Rab11a-GppNHp and the Q70 side chain is found in different positions. In the Rab11a-GTP γ S structure, S40 is closer to S25 and S42 does not interact with the γ -P oxygen. These differences indicate that the Rab11 isoforms may possess different GTP hydrolysis rates. In addition, the Switch II of inactive Rab11b presents a 3_{10} -helix (residues 69–73) that disappears upon activation. This 3_{10} -helix is not found in the Rab11a-GDP structure, which possesses a longer $\alpha 2$ helix, spanning from residue 73 to 82 α -helix 5.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Human Rab GTPases; Crystal structure; GTP binding

1. Introduction

Rab GTPases have been found in all eukaryotes, constituting the largest family of small monomeric GTPases. The human genome encodes over 60 Rab GTPases and most are ubiquitously expressed (Seabra et al., 2002). As all of the small monomeric GTPases, Rabs cycle between inactive GDP- and active GTP-bound forms and the active site regions, termed Switch I and II, undergo the larger confor-

mational changes between the two states. In the active form, they interact with effector proteins to regulate specific trafficking events such as vesicle docking, budding, motility or fusion (Stein et al., 2003). They share conserved residues related to nucleotide binding and hydrolysis and main sequence divergences lie in the carboxyl termini, which are believed to be responsible for their differences in subcellular targeting, although other regions may also be involved in proper localization (Ali et al., 2004; Chavrier et al., 1991).

The Rab11 subfamily comprises Rab11a and Rab11b, sharing 90% amino acid identity, and Rab25. This subfamily acts in recycling of proteins from the endosomes to the plasma membrane, in polarized transport in epithelial cells,

* Corresponding authors. Fax: +55 19 3512 1004.

E-mail addresses: beatriz@lnls.br (B.G. Guimarães), zanchin@lnls.br (N.I.T. Zanchin).

in the transport of molecules of the trans-Golgi network to the plasma membrane and in phagocytosis (Chen et al., 1998; Cox et al., 2000; Ullrich et al., 1996; Wang et al., 2000). The family of Rab11 interacting proteins, FIPs, comprises three classes of proteins that, in addition to the C-terminal conserved Rab11/25 binding domain (Meyers and Prekeris, 2002), contain either a C2 domain (Class I: Rip11, FIP2, and RCP) or two *EF-hand* domains and a proline rich region (Class II: FIP3/eferina, FIP4), or lack any other conserved domain (Class III: FIP1) (Prekeris, 2003). An intriguing question concerns the mechanism by which Rab GTPases generate specificity for a diverse spectrum of effectors and regulatory factors. Calorimetric studies have shown that the FIPs cannot distinguish between Rab11a and Rab11b in vitro, suggesting that they might play redundant roles in the cell (Junutula et al., 2004).

The crystal structure of Rab11a has already been determined, describing the GDP-bound form as a dimer with the Switch I and II regions involved in monomer interaction (Pasqualato et al., 2004). This finding led to the hypothesis that inactive Rab11a could dimerize in vivo and cycle between the GDP- and GTP-bound forms on the membranes without recycling to the cytosol. A crystallographic dimer involving the Switch regions has also been described for Rab9 (Wittmann and Rudolph, 2004). However, the interaction is different relative to monomer orientation. In Rab9, the C-termini of the monomers show opposite orientation whereas in Rab11a they are located on the same side of the dimer. The monomer-interacting areas of Rab11a ($\sim 2000 \text{ \AA}^2$) and Rab9 ($1200 \text{ \AA}^2$) dimers fall inside the biologically functional range of interacting areas observed for protein complexes ($1600 \pm 400 \text{ \AA}^2$) (Le Conte et al., 1999) but, so far there is no experimental evidence that Rab9 and Rab11a can dimerize in solution. Soluble dimers have been described for the Cdc42, Rac2, RhoA, and Rac1 members of the Rho family of small monomeric GTPases (Zhang and Zheng, 1998; Zhang et al., 2001). However, in these cases, the interactions occur specifically via the carboxyl-terminal region, not involving Switch I and II.

In this work, we describe Rab11b crystal structures of both inactive and active forms. Comparison of Rab11b to the previously reported Rab11a structures revealed a different oligomerization state for the inactive Rab11 form. In addition, each Rab11 isoform displays several unique interactions in the nucleotide binding site. These differences are intriguing because they share 90% sequence identity and indicate that these two isoforms may show different GTP binding or hydrolysis rates.

2. Materials and methods

2.1. Construction of expression vectors

The coding sequence of Rab11b GTPase (NCBI Accession No. XM_058232.3) was initially PCR-amplified from a human fetal brain cDNA library (Clontech) using oligonucleotides ONZ60 (5'-ACG GCA TAT GGG GAC CCG

AGA CGA CGA G-3') and ONZ61 (5'-TGG AGG ATC CGC ACG CAC GCT GGG TGG AG-3'), and inserted into the *NdeI* and *BamHI* restriction sites of the *Escherichia coli* expression vector pET28a (Novagen, Inc.), producing vector pET28-Rab11b. A second expression vector was constructed to express a truncated Rab11b form containing amino acids 8–205. For this purpose, the Rab11b cDNA cloned into pET28a was used as a template in a PCR with oligonucleotides ONZ159 (5'-GGG GGA TCC TCA GTC CGT GGT GGG CGG CAC G-3') and ONZ160 (5'-GAC CAT ATG TAC GAC TAC CTA TTC AAA GTG G-3'). The resulting PCR product was cloned into the *NdeI* and *BamHI* sites of *E. coli* vector pCYTEXP3 (Schnepppe et al., 1994), producing vector pCYTEX-Rab11b. The sequence of the expression vectors were verified by DNA sequencing analysis using an ABI Prism 377 DNA sequence analyzer (Applied Biosystems).

2.2. Expression and purification of Rab11b

Rab11b was expressed in *E. coli* strain DH5 α transformed with vector pCYTEX-Rab11b. Cells were grown at 30 °C in LB medium containing ampicillin ($50 \mu\text{g ml}^{-1}$) to an optical density at 600 nm (OD_{600}) of 0.8–1.0 and protein expression was induced by heat-shock at 42 °C for 2 h. The frozen bacterial pellet from 2 L cultures was lysed in 50 ml of buffer A (20 mM Tris-HCl, pH 8.5, 20 mM NaCl, 0.1 mM DTT, 1 mM MgCl_2) containing 0.5 mM PMSF and $50 \mu\text{g ml}^{-1}$ lysozyme for 1 h on ice. Cell extracts were isolated by sonication in a Branson sonifier (Branson Ultrasonics, Co) and centrifugation at 20000g for 30 min at 4 °C. For Rab11b purification, the extract was loaded onto a 45 cm^3 DEAE-Sepharose FF ion-exchange column equilibrated in buffer A. Rab11b was eluted with a 450 ml (10 column volumes, CV) linear gradient from 20 to 200 mM NaCl in buffer A. Fractions enriched with Rab11b were pooled, dialyzed against buffer A and loaded onto a 16 cm^3 Q-Sepharose HP column, which was eluted using a 160 ml (10 CV) gradient from 20 to 300 mM NaCl in buffer A. Rab11b eluted from this column at approximately 100 mM NaCl. Further purification was performed with a Superdex 75 16/60 column using a buffer containing 10 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 1 mM MgCl_2 . Approximately 60 mg of pure Rab11b were obtained with this expression system. All chromatographic procedures were performed using an ÄKTA-FPLC system and columns purchased from GE-Healthcare (former Amersham Biosciences). Protein concentration was estimated from direct absorbance at 280 nm in guanidine buffer (7.2 M guanidine-HCl, 25 mM sodium phosphate buffer, pH 6.55) considering an extinction molar coefficient of $\epsilon_{280} = 21620 \text{ M}^{-1} \text{ cm}^{-1}$ for Rab11b.

2.3. Nucleotide exchange

The GDP molecule that co-purifies with recombinant Rab11b and Rab11a was exchanged against GppNHp by

incubating the purified proteins with 5 mM GppNHp, 5 mM MgCl_2 and 5 U of calf intestinal alkaline phosphatase per milligram of protein at room temperature for 16 h. The excess nucleotide and alkaline phosphatase were subsequently removed by gel-filtration with a Superdex 75 16/60 column using buffer containing 10 mM Tris–HCl, pH 7.4, 150 mM NaCl, and 1 mM MgCl_2 .

2.4. Gel-filtration analysis of Rab11b

Gel-filtration experiments were performed on a Superdex 75 16/60 column. The molecular weight markers used were bovine serum albumin (67 kDa, 7 mg ml^{−1}); ovalbumin (43 kDa, 7 mg ml^{−1}); chymotrypsinogen (25 kDa, 3 mg ml^{−1}), and ribonuclease A (14 kDa, 10 mg ml^{−1}). Both the molecular weight markers and the sample were prepared in buffer containing 10 mM Tris–HCl, pH 7.4, 150 mM NaCl, and 1 mM MgCl_2 . Additional analyses were performed with the same buffer containing also 0.1% Tween 20 or 9 mM deoxycholate.

2.5. Crystallization and diffraction data collection

Crystallization systems and reagents were purchased from Hampton Research and Jena Bioscience and proteins were crystallized by the hanging-drop vapor-diffusion method using 24-well plates at 18°C. Rab11b was crystallized in its inactive (Rab11b-GDP) and active (Rab11b-GppNHp) forms. First crystals of the Rab11b-GDP complex were grown in drops containing equal volumes (2 μ l) of protein sample (30 mg ml^{−1} in 10 mM Tris–HCl, pH 7.4, 0.1 mM MgCl_2), and reservoir solution (0.1 M Tris–HCl, pH 8.5, 30% PEG 4000, and 0.2 M sodium acetate), equilibrated against 300 μ l of reservoir solution. Withdrawal of sodium acetate from the crystallization solution improved crystal morphology. Crystals were further optimized by testing PEG 4000 concentrations ranging from 25 to 32% and pH from 8.0 to 9.0. Crystals for data collection (reaching up to 500 μ m at the maximum dimension) were obtained in the condition containing 30% PEG 4000 and 0.1 M Tris–HCl, pH 8.8. Rab11b-GppNHp did not crystallize under the same conditions as Rab11b-GDP. Therefore, a crystallization screen of the Rab11b-GppNHp complex was performed using 2 μ l of protein at 68 mg ml^{−1} in 10 mM Tris–HCl, pH 7.4, and 0.1 mM MgCl_2 , which were mixed with equal volumes of the reservoir solutions and equilibrated against 300 μ l of reservoir solutions. Crystals used for data collection (~100 μ m at the maximum dimension) were formed in the condition 41 of the Hampton Research Crystal Screen II (0.1 M Tris–HCl, pH 8.5, 10 mM NiCl_2 , and 1 M Li_2SO_4).

Crystals were transferred directly into a cryoprotectant solution which contained the respective reservoir solutions used to crystallize inactive and active Rab11b supplemented with 25% glycerol. Crystals were flash-cooled under a nitrogen stream at 100 K and complete data sets were collected at the protein crystallography beam line D03B-MX1

of the Brazilian Synchrotron Light Laboratory, using a MAR165 CCD detector at a wavelength of 1.430 Å. Data were processed using the programs MOSFLM (Leslie, 1992) and SCALA (Blessing, 1995; Evans, 1993; Kabsch, 1988) from the CCP4 package (Collaborative Computational Project 4, 1994). Crystals of inactive Rab11b (Rab11b-GDP) diffract to 1.55 Å resolution and belong to the space group $P2_12_12_1$ whereas crystals of Rab11b in complex with the non-hydrolyzable GTP analogue GppNHp diffracted to 1.95 Å resolution and belong to the space group $I2_12_12_1$. A summary of the data collection statistics is shown in Table 1.

2.6. Structure solution and refinement

The crystal structures were solved by molecular replacement methods using the program AMoRe (Navaza, 1994) as implemented in the CCP4 program suite. The structure of Rab11a (PDB code 1OIV, Pasqualato et al., 2004) was used as the search model to determine the structure of Rab11b-GDP. The Rab11b-GDP structure, excluding the Switch I and II regions, was used as the starting model to determine the structure of Rab11b-GppNHp. The structures were refined using Refmac5 (Murshudov et al., 1997). Refinement cycles were alternated with visual inspection of the electron density maps and model rebuilding with the program O (Jones and Kjeldgaard, 1997). During the final cycles water molecules were introduced using the program ARP/WARP (Lamzin and Wilson, 1993). The final models present an R_{factor} of 0.186 ($R_{\text{free}} = 0.238$) and R_{factor} of 0.163 ($R_{\text{free}} = 0.205$) for Rab11b-GDP and Rab11b-GppNHp respectively, with good overall stereochemistry. As defined by the program PROCHECK (Laskowski et al., 1993) all non-glycine and non-proline residues fall in the most favored or additionally allowed regions of the Ramachandran plot. Further details of refinement are presented in Table 2. Rab11b-GDP and Rab11b-GppNHp atomic coordinates were deposited in the Protein Data Bank under the codes 2F9L and 2F9M, respectively.

Table 1

Data collection statistics (values in parenthesis are for the outer resolution shell)

	Rab11b-GDP	Rab11b-GppNHp
Beam line	LNLS D03B	LNLS D03B
Wavelength (Å)	1.430	1.430
Space group	$P2_12_12_1$	$I2_12_12_1$
Unit-cell parameters (Å)	$a = 45.79$ $b = 52.22$ $c = 59.35$	$a = 66.61$ $b = 85.41$ $c = 86.87$
Resolution limits (Å)	39.22–1.55 (1.63–1.55)	60.86–1.95 (2.06–1.95)
Total observations	141 959	168 903
Unique reflections	21 134	18 222
Multiplicity	6.7 (6.4)	9.2 (9.0)
Completeness (%)	99.8 (99.5)	99.8 (100.0)
Rsym (%)	6.2 (37.7)	9.0 (33.1)
Mean I/ σ (I)	18.4 (5.0)	19.3 (6.6)

Table 2
Refinement statistics

	Rab11b-GDP	Rab11b-GppNHp
Resolution range (Å)	39.22–1.55	60.86–1.95
No. of reflections	20 119	17 437
No. of protein/ligand/water atoms	1341/29/185	1456/33/286
R_{factor} (%)	18.6	16.3
R_{free} (%)	23.8	20.5
RMS deviations from ideality		
Bond lengths (Å)	0.014	0.009
Bond angles (degrees)	1.720	1.263
Ramachandran plot (%)		
Most favored regions	92.4	94.7
Additional allowed regions	7.6	5.3
Average B -factor (Å ²)		
Main chain/side chain	15.72/17.58	14.03/16.36
Ligand/solvent	10.82/30.83	10.43/31.58

3. Results and discussion

3.1. Overall Rab11b structure

Attempts to crystallize the full-length Rab11b protein have failed and, since the N- and C-terminal flexible regions of small GTPases are known to interfere with crystallization, a construct was made to remove seven residues from the N-terminal and 13 from the C-terminal regions of Rab11b, so that the variant used in this work includes amino acids 8 to 205 with an additional methionine at the N-terminus. The crystal structures of GDP- and GppNHp-bound forms were refined at 1.55 and 1.95 Å resolution, respectively. In both cases, Rab11b crystallized as a monomer in the asymmetric unit. The final atomic model of inactive Rab11b includes residues 7–182, a bound GDP molecule and a magnesium ion. The quality of the electron density maps did not allow modeling of residues 39–41 of the Switch I, which is consistent with the fact that Rab

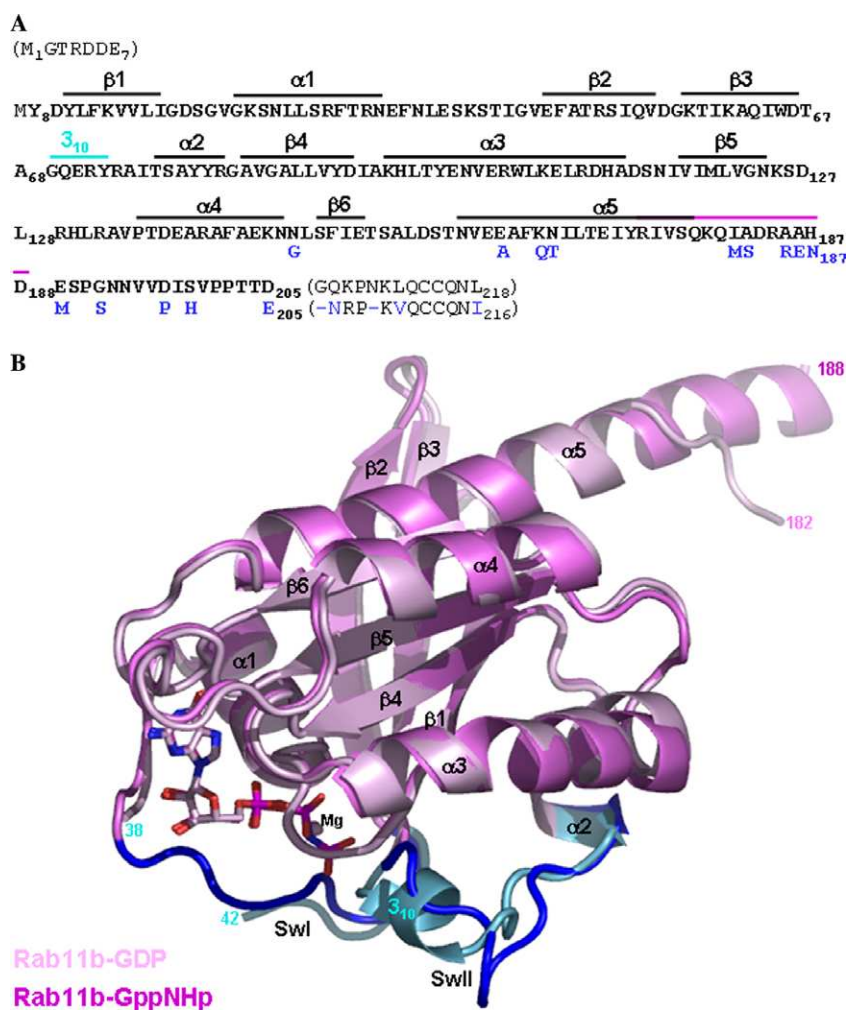


Fig. 1. Overall structure of Rab11b. (A) Sequence of Rab11b showing the secondary structure elements of the inactive and active forms. The amino acid substitutions in Rab11a are indicated in below the Rab11b sequence. The $\beta 3_{10}$ -helix present only in inactive Rab11b is indicated in cyan. α -Helix 5 is longer in the active form (extension indicated in magenta). (B) Superposition of inactive and active Rab11b structures. The Switch I and II regions are shown in blue and Rab11b-GDP is represented in lighter colors.

inactive forms usually show a mobile Switch I region. The final model of the Rab11b-GppNHp complex includes residues 7–188. The structure of monomeric Rab11b shows the typical Ras-like, small GTPase fold with a six stranded β -sheet core ($\beta 1$ – $\beta 6$) surrounded by five α -helices ($\alpha 1$ – $\alpha 5$) (Fig. 1). Superposition of Rab11b-GDP and Rab11b-GppNHp structures results in an overall RMS deviation of 0.79 Å (165 C α aligned).

Major conformational differences in the two forms of Rab11b involve helices $\alpha 2$ and $\alpha 5$. In the GDP-bound complex, amino acids 69–73 form a 3_{10} -helix followed by an $\alpha 2$ (residues 77–82) smaller than the one present in other Rab structures. In the GppNHp-bound complex, the 3_{10} -helix disappears whereas $\alpha 2$ remains (Fig. 1). This change in the $\alpha 2$ region upon activation is observed only for the Rab11 isoforms. Other Rab proteins contain a conserved and longer α -helix in the Switch II region (Fig. 2). As for $\alpha 5$, while in the Rab11b-GDP complex this α -helix includes residues 161–178, Rab11b-GppNHp shows an $\alpha 5$ spanning from amino acid 161 to 188 (Fig. 1), indicating a stabilization of the 179–188 region following binding to GppNHp.

The crystal structure of the Rab11a isoform bound to GDP and of a Rab11a Q70L variant bound to GTP γ S was described by Pasqualato et al. (2004). These authors have reported that inactive Rab11a crystallized as a dimer in the asymmetric unit. By contrast, both the GDP- and the GppNHp-bound Rab11b structures show a single crystallographically independent monomer. Superposition of Rab11b monomers and Rab11a-GDP and Rab11a(Q70L)-GTP γ S results in overall RMS deviation of 0.90 Å (153 C α aligned) and 0.55 Å (166 C α aligned) for the inactive (Fig. 3) and active forms, respectively. A second crystal structure of the active Rab11a bound to GppNHp was reported by Eathiraj et al. (2005). Superpo-

sition of Rab11a and Rab11b bound to GppNHp results in overall RMS deviation of 0.44 (168 C α aligned). Despite the similarity, the overall alignment for the isoforms shows some differences in the Switch regions for the Rab11-GDP complexes. As mentioned above, the Rab11b-GDP structure lacks residues 39–41 of Switch I in contrast to inactive Rab11a structure in which the entire region could be modeled. The conformational differences between the Switch regions of inactive Rab11a and Rab11b could be in part explained by the dimerization of Rab11a-GDP in the crystal, which is consistent with the proposal that in inactive Rabs the Switch regions are either disordered or influenced by crystal contacts (Eathiraj et al., 2005). The Switch II of inactive Rab11b presents a 3_{10} -helix (residues 69–73) that disappears upon activation. This 3_{10} -helix is not found in the Rab11a-GDP structure, which possesses a longer $\alpha 2$ helix, spanning from residue 73 to 82. (Fig. 3A).

3.2. Analysis of the oligomerization state of Rab11

Inactive Rab9 and Rab11a have been crystallized as dimers (Pasqualato et al., 2004; Wittmann and Rudolph, 2004) and both dimers show large interaction interfaces (2000 Å² in Rab11a and 1200 Å² in Rab9) and it has been speculated that Rab proteins can form dimers in vivo. Dimerization of the Rho family GTPases Cdc42, Rac2, and RhoA in solution has been described (Zhang and Zheng, 1998). In the case of Cdc42 and Rac2, it was also shown that dimerization leads to an increase of the intrinsic GTPase activity (Zhang and Zheng, 1998). It is important to point out that dimerization of Rho family members takes place via the C-terminal region, not involving the Switch I and II regions. The Rab11a-GDP crystal structure showed a dimer in the asymmetric unit, but assays to detect Rab11a dimers in solution have not been successful (Pasqualato et al., 2004). The Rab11a dimer interface buries large fractions of Switch I and Switch II regions, which are normally involved in the interactions of Rab proteins with their partners. In this work, size exclusion chromatography was performed to determine the molecular weight of inactive Rab11b under various conditions, including in the presence of detergents intended to simulate a membrane-like environment, as proposed previously to explain Rab11a dimer formation (Pasqualato et al., 2004), but no dimer could be detected (data not shown). Both inactive and active Rab11b have crystallized as monomers in the asymmetric unit and analysis of the neighbors in the crystal lattice did not reveal any possible dimer. Some Rab11b sequences in the database contain an arginine residue at position 75 of the Switch II region but the isoform used in this work contains an alanine in this position, showing amino acid identity to Rab11a up to residue 147 (Fig. 1A). Structural alignment of inactive forms of Rab11a and Rab11b shows that the amino acid substitutions are far away from the Rab11a dimer interface making the difference in their quaternary structures unexpected (Fig. 3B). Furthermore, analysis of the Rab11a dimer using the PISA webserver

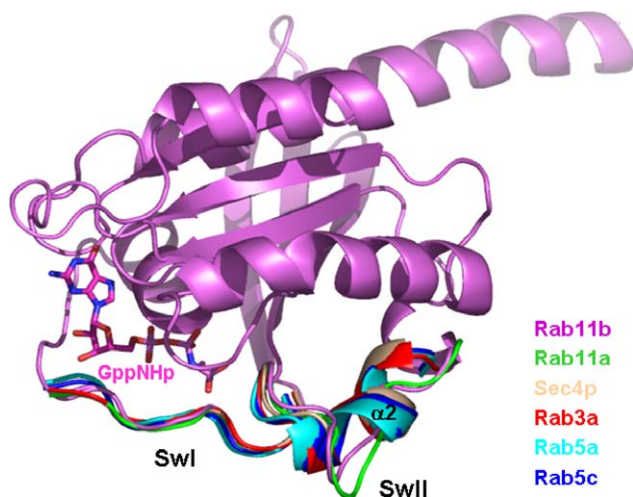


Fig. 2. Structural comparison of the Switch I and II regions of active Rabs. Rab11a and Rab11b possess closely related Switch II and differ from other Rabs that contain helical Switch II. In the Switch I region, the differences are not significant. For best visualization, only the core of Rab11b is shown.

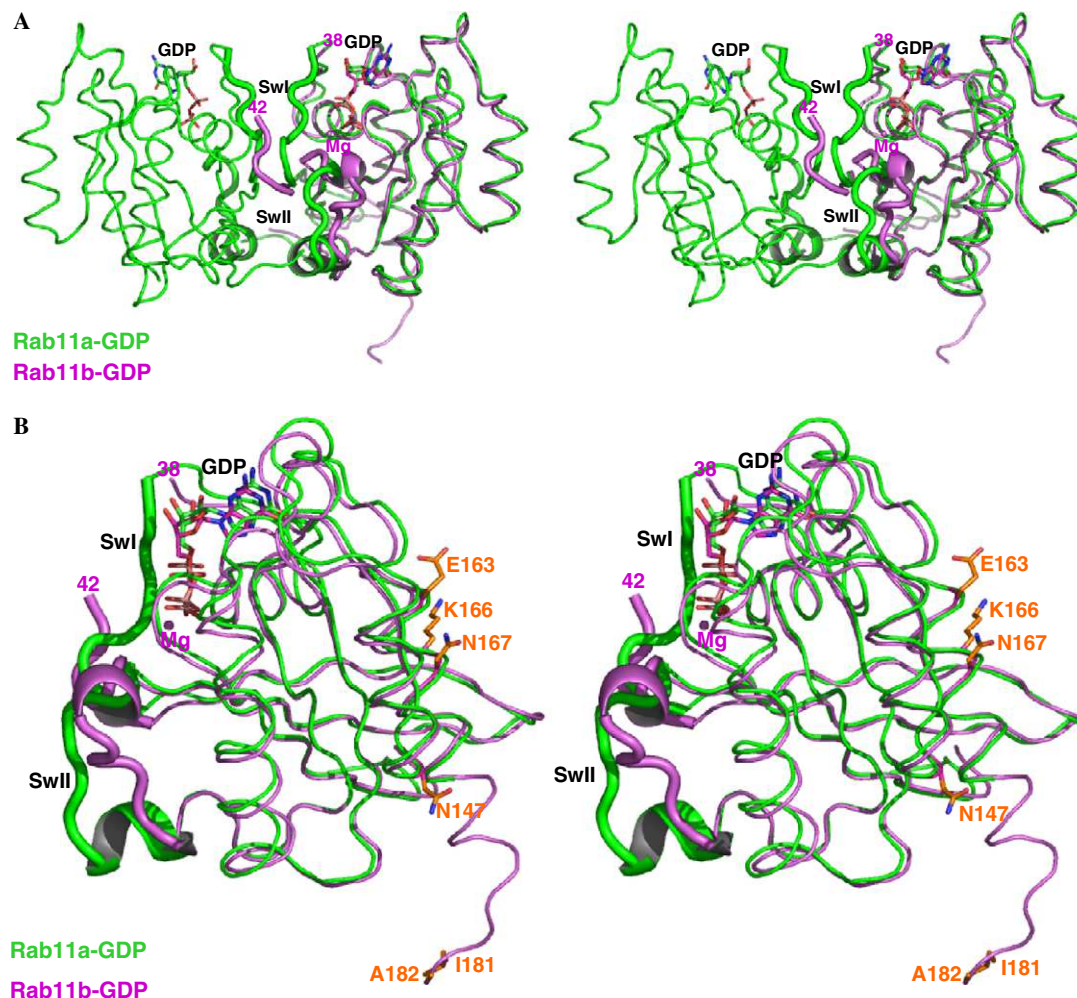


Fig. 3. Stereo view of C- α superposition of Rab11a (green) and Rab11b (magenta) GTPases. The Switch regions, that undergo large conformational changes during the GDP/GTP cycle, are highlighted. (A) In its inactive form, Rab11b crystallized as a monomer in complex with a GDP molecule and a magnesium ion. Rab11a-GDP crystallized as a dimer and lacks the bound magnesium ion at the nucleotide binding site. Rab11b-GDP structure lacks residues 39–41 from Switch I. (B) Rab11b side chains corresponding to amino acid substitutions between the isoforms are shown. Note that the substitutions are far from the Rab11a dimer interface. A single monomer of Rab11a is represented.

(http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html), that calculates protein assemblies from crystal structures based on general principles of chemical thermodynamics (Krissinel and Henrick, 2005), revealed that the theoretical ΔG values for association and dissociation of the Rab11a monomers are about -9 and 3 kcal mol^{-1} , respectively, which indicates a weak binding, despite the large surface area buried in the dimers. Based on these analyses, we are confident that the monomer described in this work for inactive Rab11b represents a biologically functional intermediate of the protein. The crystal structure alone did not suggest a mechanism for Rab dimerization neither revealed a reason why the two isoforms would present different quaternary structures.

3.3. Interactions in the nucleotide binding site

The nucleotide binding site of the monomeric Rab11b-GDP complex shows interactions between the phosphate

groups of GDP and the conserved amino acids of the P-loop (residues 18–25) that are typical of small GTPases. The binding site also displays a disordered Switch I and a magnesium ion coordinated by four water molecules, the phosphate β of GDP (β -P) and the invariant serine 25 (Fig. 4A). The Rab11a-GDP dimer, in contrast, lacks the bound magnesium ion at the nucleotide binding site (Pasqualato et al., 2004). In the Rab11a-GDP crystal, the Switch regions assume an unusually ordered conformation, probably due to stabilizing contacts present in the large dimeric interface. The unusually ordered Rab11a-GDP Switch I region and the resulting interaction of S40 with the α -P and the ribose of GDP contribute to maintain the nucleotide tightly bound in the absence of magnesium (Pasqualato et al., 2004). In addition, the invariant Mg-binding S25 from the P-loop shows an unusual conformation and interacts with α -P instead of β -P of the GDP (Fig. 4A). The amino acid sequence of the Rab11b variant used in this work is identical to Rab11a up to amino acid

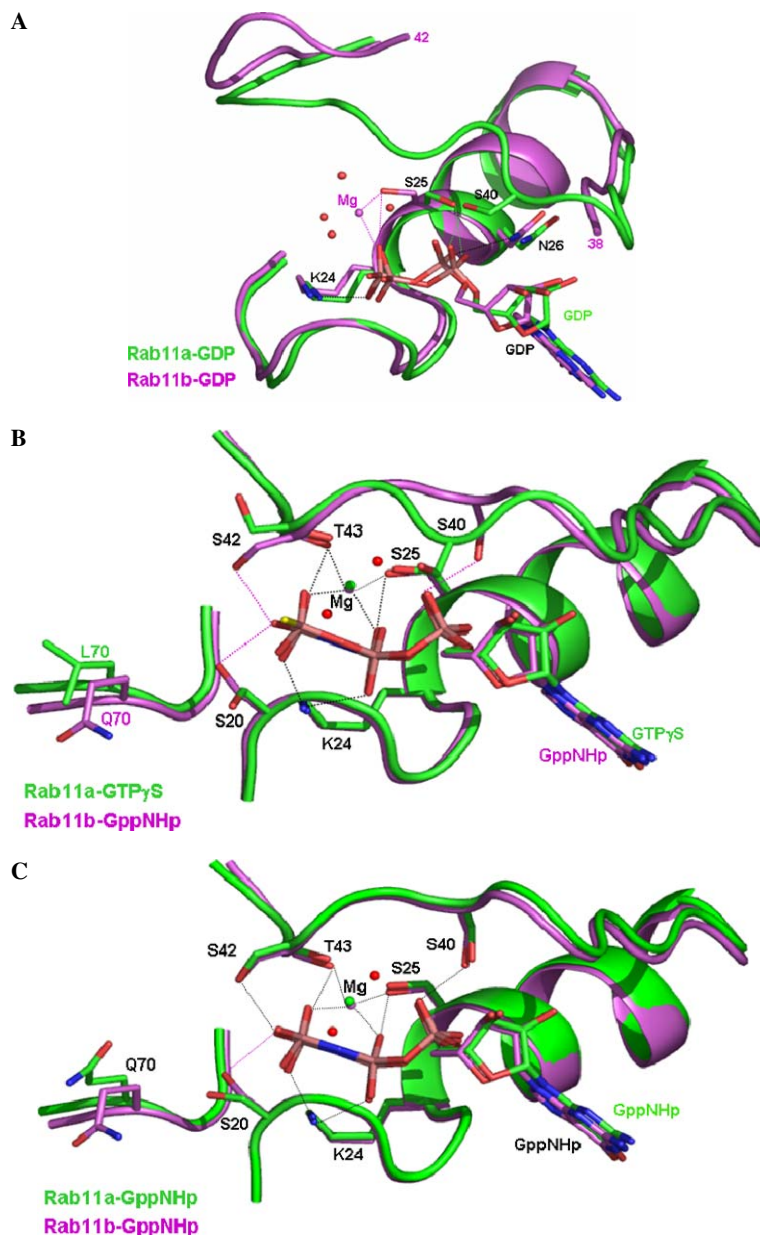


Fig. 4. Superposition of the nucleotide binding sites of Rab11a (green) and Rab11b (magenta) structures. The side chains of residues involved in nucleotide stabilization are shown. Magnesium ions found in Rab11a and Rab11b structures are shown as green and magenta spheres, respectively, and their coordinating water molecules as red spheres. Interactions are represented as green and magenta dotted lines for Rab11a and Rab11b structures, respectively. Interactions that are present in both structures are represented only for Rab11b in black lines. (A) Nucleotide interactions of inactive Rab11a and Rab11b. The ordered Rab11a-GDP Switch I region and the resulting interaction of S40 with the α -P and the ribose of GDP contribute to maintain the nucleotide tightly bound in the absence of magnesium. The invariant Mg-binding S25 from the P-loop of inactive Rab11a shows an unusual conformation and interacts with α -P instead of β -P of GDP. (B) Nucleotide interactions of active Rab11a (Rab11a-Q70L-GTP γ S) and Rab11b (Rab11b-GppNHp). In contrast with active Rab11b and other active Rab structures, neither S20 from the P-loop nor S42 from Switch I of Rab11a(Q70L)-GTP γ S interact with the γ -P oxygen. S40 is also displaced. The Q70L mutation of active Rab11a is also shown. (C) Nucleotide interactions of active Rab11a (Rab11a-GppNHp) and Rab11b (Rab11b-GppNHp). S40 and S42 show a good overlapping in both structures and S20 does not interact with the γ -P oxygen of Rab11a-GppNHp. Q70 side chain is found in different positions.

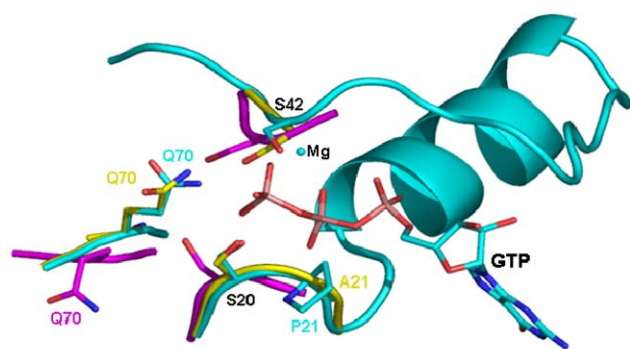
147, which includes the Switch regions (Fig. 1A), making the structural differences between these two GTPases unexpected.

Although superposition of the Rab11a(Q70L)-GTP γ S and Rab11b-GppNHp structures shows a close alignment with a small displacement in the Switch regions, several key

differences in side chain conformation of the residues involved in GTP hydrolysis are observed. In contrast to active Rab11b and other Rab structures, neither S20 from the P-loop nor S42 from the Switch I of active Rab11a(Q70L)-GTP γ S are in position to interact with the γ -P oxygen (Fig. 4B). These residues are associated with the

intrinsic GTPase hydrolytic rate, and the conformational differences suggest that the Rab11 isoforms could differ in hydrolysis kinetics. S40 is also displaced in the Rab11a(Q70L)-GTP γ S structure, but it is not known if it can interfere with the hydrolysis process (Fig. 4B). The Q70L mutation introduced in Rab11a was intended to generate a stable GTP γ S-bound form, however the mutant showed GTP hydrolysis rates similar to the wild-type protein (Pasqualato et al., 2004). Furthermore, it is not clear what would be the effect of having a sulfur atom bound to γ -P instead of an oxygen. Interestingly, both S40 and S42 show a good overlapping when the Rab11a-GppNHp (Eathiraj et al., 2005) and Rab11b-GppNHp structures are compared.

Q70 is highly conserved and has been a frequent target for mutagenesis to generate constitutively active GTPases and it is not possible to predict to which extent this mutation influences S20 and S42 positions in the Rab11a(Q70L)-GTP γ S structure. In this context, it is important to compare the interactions of the nucleotide binding site of Rab5a that was described for the active wild-type and A21P hydrolysis deficient mutant proteins (Zhu et al., 2003, 2004). Mutations in the position equivalent to S20 in the Ras GTPase have been associated to a slower GTP hydrolysis rate and increased biological activity (Seeburg et al., 1984). The A21P substitution in Rab5a resulted in a hydrolysis deficient mutant (Zhu et al., 2003, 2004) and did not affect the correct position of S20, but affected the side chain position of S42, compromising its interaction with the γ -P of GTP (Fig. 5). Interestingly, a significant conformational difference between active Rab11b and Rab5a crystal structures involves Q70 (Fig. 5), which does not interact with the γ -P of GppNHp in Rab11b but interacts with the γ -P of GTP of Rab5a (Zhu et al., 2003).



Rab5a (A21P)-GTP
Rab5a-GppNHp
Rab11b-GppNHp

Fig. 5. Active site of the hydrolysis deficient Rab5a A21P mutant (PDB code 1N6L) crystallized in complex with GTP, showing the side chains of residues probably involved in hydrolysis (cyan). Residues at equivalent positions in wild-type Rab5a-GppNHp and Rab11b-GppNHp are colored yellow and pink, respectively. Residues are numbered according to Rab11b amino acid positions. Serine 20 and 42 are usually mentioned as determinants of hydrolytic rate. The A21P mutation promoted the loss of interaction between the side chain of S42 and the γ -P of GTP in Rab5a. Q70 also shows significant conformational change, which does not interact with the γ -P of GTP in Rab11b.

Hydrolysis deficient mutants are usually crystallized bound to GTP, not requiring a non-hydrolyzable GTP analog. This is not the case for the Rab11a Q70L substitution which did not affect its GTP hydrolysis rate (Pasqualato et al., 2004). Therefore, it is expected that the Q70L mutation in Rab11a did not influence the side chain position of S20 and S42. A comparison of Rab11b-GppNHp and Rab11a-GppNHp (Eathiraj et al., 2005) shows that the Q70 side chain is found in different positions (Fig. 4C) and in both cases does not interact with the γ -P oxygen, as opposite to Rab5a. An effective reduction of the GTP hydrolysis rate for Rab11a was achieved using a S20V mutant (Chen and Wandinger-Ness, 2001). Interestingly, S20 does not interact with γ -P oxygen of both Rab11a(Q70L)-GTP γ S and Rab11a-GppNHp but interacts with this oxygen of Rab11b-GppNHp, Rab5a(A21P)-GTP, and Rab5a-GppNHp (Figs. 4 and 5).

In conclusion, the differences observed in the side chains of S20, S40, and S42 of active Rab11b relative to Rab11a indicate that these closely related isoforms may present different GTP hydrolysis rates. In addition, the conformational “flexibility” of Q70 may also be a factor affecting the process of GTP hydrolysis among the different Rab proteins.

Acknowledgments

This work was supported by Grant 00/02788-4 and the SMolBNet and CEPID/CBME programs from the Foundation for Research Support of the State of São Paulo (FAPESP). S.M.N.S. and F.R.G.C. are recipients of FAPESP pre-doctoral fellowships. We thank Tereza C. Lima Silva for technical support, Zildene G. Correa and Luciana R. Camillo for DNA sequencing.

References

- Ali, B.R., Wasmeier, C., Lamoreux, L., Strom, M., Seabra, M.C., 2004. Multiple regions contribute to membrane targeting of Rab GTPases. *J. Cell Sci.* 117, 6401–6412.
- Blessing, R.H., 1995. An empirical correction for absorption anisotropy. *Acta Crystallogr.* A51, 33–38.
- Chavrier, P., Gorvel, J.P., Stelzer, E., Simons, K., Gruenberg, J., Zerial, M., 1991. Hypervariable C-terminal domain of Rab proteins acts as a targeting signal. *Nature* 353, 769–772.
- Chen, W., Feng, Y., Chen, D., Wandinger-Ness, A., 1998. Rab11 is required for trans-Golgi network-to-plasma membrane transport and a preferential target for GDP dissociation inhibitor. *Mol. Biol. Cell* 9 (11), 3241–3257.
- Chen, W., Wandinger-Ness, A., 2001. Expression and functional analyses of Rab8 and Rab11a in exocytic transport from trans-Golgi network. *Methods Enzymol.* 329, 165–175.
- Collaborative Computational Project, Number 4, 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* D 50: 760–763.
- Cox, D., Lee, D.J., Dale, B.M., Calafat, J., Greenberg, S., 2000. A Rab11-containing rapidly recycling compartment in macrophages that promotes phagocytosis. *Proc. Natl. Acad. Sci. USA* 97, 680–685.
- Eathiraj, S., Pan, X., Ritacco, C., Lambright, D.G., 2005. Structural basis of family-wide Rab GTPase recognition by rabenosyn-5. *Nature* 436 (7049), 415–419.

- Evans, P.R., 1993. Data reduction, Proceedings of CCP4 Study Weekend on Data Collection & Processing, pp. 114–122.
- Jones, T.A., Kjeldgaard, M., 1997. Electron-density map interpretation. *Methods Enzymol.* 277, 173–208.
- Junutula, J.R., Schonteich, E., Wilson, G.M., Peden, A.A., Scheller, R.H., Prekeris, R., 2004. Molecular characterization of Rab11 interactions with the members of family of Rab11-interacting proteins (FIPs). *J. Biol. Chem.* 279 (32), 33430–33437.
- Kabsch, W., 1988. Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Crystallogr.* 21, 916–924.
- Krissinel, E., Henrick, K., 2005. Detection of Protein Assemblies in Crystals. In: Berthold, M.R. (Ed.), *CompLife 2005*, LNBI 3695 Springer-Verlag, Berlin, Heidelberg, pp. 163–174.
- Lamzin, V.S., Wilson, K.S., 1993. Automated refinement of protein models. *Acta Crystallogr. D Biol. Crystallogr.* 49 (1), 129–147.
- Laskowski, R.A., Moss, D.S., Thornton, J.M., 1993. Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* 231 (4), 1049–1067.
- Le Conte, L., Chothia, C., Janin, J., 1999. The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* 285 (5), 2177–2198.
- Leslie, A.G.W., 1992. Joint CCP4 and ESF-EACMB. In: *Newsletter on Protein Crystallography*, vol. 26. Daresbury Laboratory, Warrington, UK.
- Meyers, J.M., Prekeris, R., 2002. Formation of mutually exclusive Rab11 complexes with members of the family of Rab11-interacting proteins regulates Rab11 endocytic targeting and function. *J. Biol. Chem.* 277 (50), 49003–49010.
- Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53 (3), 240–255.
- Navaza, J., 1994. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* 50, 157–163.
- Pasqualato, S., Senic-Matuglia, F., Renault, L., Goud, B., Salamero, J., Cherfils, J., 2004. The structural GDP/GTP cycle of Rab11 reveals a novel interface involved in the dynamics of recycling endosomes. *J. Biol. Chem.* 279 (12), 11480–11488.
- Prekeris, R., 2003. Rabs, Rips, FIPs and endocytic membrane traffic. *Sci. World J.* 3, 870–880.
- Schneppe, B., Eichner, W., McCarthy, J.E., 1994. Translational regulation of a recombinant operon containing human platelet-derived growth factor (PDGF)-encoding genes in *Escherichia coli*: genetic titration of the peptide chains of the heterodimer AB. *Gene* 143 (2), 201–209.
- Seabra, M.C., Mules, E.H., Hume, A.N., 2002. Rab GTPases, intracellular traffic and disease. *Trends Mol. Med.* 8, 23–30.
- Seeburg, P.H., Colby, W.W., Capon, D.J., Goeddel, D.V., Levinson, A.D., 1984. Biological properties of human c-Ha-ras1 genes mutated at codon 12. *Nature* 312 (5989), 71–75.
- Stein, M.-P., Dong, J., Wandinger-Ness, A., 2003. Rab proteins and endocytic trafficking: potential targets for therapeutic interventions. *Adv. Drug Deliv. Rev.* 55, 1421–1437.
- Ullrich, O., Reinsch, S., Urbe, S., Zerial, M., Parton, R.G., 1996. RAB11 regulates recycling through the pericentriolar recycling endosome. *J. Cell Biol.* 135 (4), 913–924.
- Wang, X., Kumar, R., Navarre, J., Casanova, J.E., Goldenring, J.R., 2000. Regulation of vesicle trafficking in madin-darby canine kidney cells by RAB11A and RAB25. *J. Biol. Chem.* 275 (37), 29138–29146.
- Wittmann, J.G., Rudolph, M.G., 2004. Crystal structure of Rab9 complexed to GDP reveals a dimer with an active conformation of switch II. *FEBS Lett.* 568 (1–3), 23–29.
- Zhang, B., Zheng, Y., 1998. Negative regulation of Rho family GTPases Cdc42 and Rac2 by homodimer formation. *J. Biol. Chem.* 273 (40), 25728–25733.
- Zhang, B., Gao, Y., Moon, S.Y., Zhang, Y., Zheng, Y., 2001. Oligomerization of Rac1 GTPase mediated by the carboxyl-terminal polybasic domain. *J. Biol. Chem.* 276 (12), 8958–8967.
- Zhu, G., Liu, J., Terzyan, S., Zhai, P., Li, G., Zhang, X.C., 2003. High resolution crystal structures of human Rab5a and five mutants with substitutions in the catalytically important phosphate-binding loop. *J. Biol. Chem.* 278 (4), 2452–2460.
- Zhu, G., Zhai, P., Liu, J., Terzyan, S., Li, G., Zhang, X.C., 2004. Structural basis of Rab5–Rabaptin5 interaction in endocytosis. *Nat. Struct. Mol. Biol.* 11 (10), 975–983.